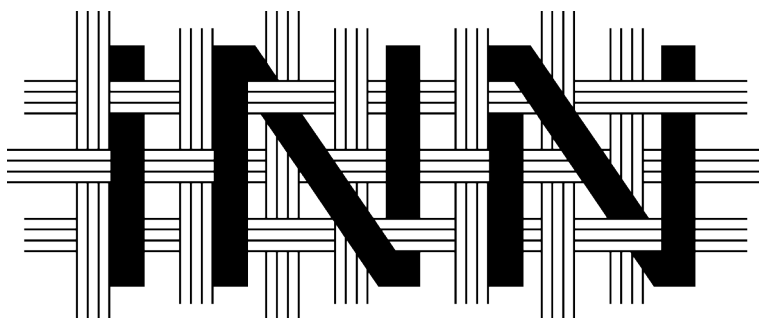


# **Kvantitativ metode og statistikk (IDR4000)**

Kandidatnummer: 505

Antall ord: 12874



## **Høgskolen i Innlandet**

2024-11-22

# Table of contents

<b>Forord</b>	<b>4</b>
<b>1 Reliabilitetstesting</b>	<b>5</b>
1.1 Introduksjon . . . . .	5
1.2 Metode . . . . .	6
1.3 Resultat . . . . .	6
1.3.1 Korrelasjon mellom $\dot{V}O_{2maks}$ og Wattmaks per kg . . . . .	7
1.4 Diskusjon . . . . .	7
<b>2 Regresjonsmodeller og prediksjoner</b>	<b>8</b>
2.1 Introduksjon . . . . .	8
2.2 Part 1 - Lactate thresholds . . . . .	8
2.2.1 Metode . . . . .	8
2.2.2 Resultat . . . . .	9
2.2.3 Diskusjon . . . . .	9
2.3 Part 2 - Predicting sizes of DNA fragments . . . . .	10
2.3.1 Metode . . . . .	10
2.3.2 Resultat . . . . .	10
2.3.3 Diskusjon . . . . .	11
2.4 Part 3 - Interpreting a regression table . . . . .	12
2.4.1 Metode . . . . .	12
2.4.2 Resultat . . . . .	12
2.4.3 Diskusjon . . . . .	13
<b>3 Å trekke slutninger fra statistiske modeller og statistisk styrke</b>	<b>14</b>
3.1 Spørsmål og svar . . . . .	14
3.1.1 Estimate . . . . .	14
3.1.2 m1 vs m2 . . . . .	14
3.1.3 Shaded areas . . . . .	15
3.1.4 Standard deviation of <b>estimate</b> and avg. <b>se</b> for each study. . . . .	15
3.1.5 P-value histogram . . . . .	15
3.1.6 Antall studier med statistisk signifikans . . . . .	16
3.1.7 Power of a one-sample t-test . . . . .	16
3.1.8 Med signifikansnivå på 0.05 hvor mange studier gir “falsk positiv” ved gjennomføring av mange repeterte studier? . . . . .	16
3.2 Code-chunks fra topp til bunn . . . . .	17
<b>4 Studiedesign</b>	<b>21</b>
4.1 Introduksjon . . . . .	21
4.2 Metode . . . . .	21
4.3 Resultat . . . . .	21

4.4	Diskusjon . . . . .	22
4.5	Konklusjon . . . . .	22
<b>5</b>	<b>Analysering av eksperimenter med gjentatte målinger</b>	<b>24</b>
5.1	Introduksjon . . . . .	24
5.2	Metode . . . . .	24
5.2.1	Forsøkspersoner og studiedesign . . . . .	24
5.2.2	Treningsprotokoll . . . . .	25
5.2.3	Målinger av muskelstyrke og hypertrofi . . . . .	25
5.2.4	Dataanalyser og statistikk . . . . .	26
5.3	Resultater . . . . .	26
5.4	Diskusjon . . . . .	27
<b>6</b>	<b>Vitenskapsteori</b>	<b>28</b>
<b>7</b>	<b>Lab-rapport</b>	<b>30</b>
7.1	Introduksjon . . . . .	30
7.2	Metode . . . . .	30
7.2.1	Databehandling . . . . .	32
7.3	Diskusjon . . . . .	33
7.3.1	Konklusjon . . . . .	33
	<b>Referanser</b>	<b>34</b>

# Forord

Denne rapporten inneholder ulike arbeidskrav som er gjennomført gjennom høsten 2024.

Arbeidskravene er satt sammen og danner nå mappeeksamen i Kvantitativ metode og statistikk (IDR4000).

Takk til Ole for mye godt samarbeid underveis, både med kode og som selskap under ferdigstillingen av eksamen.

Lenke som gjør det mulig å etterprøve koder: <https://github.com/Eskilstrand/Mappeeksamen.git>

# 1 Reliabilitetstesting

## 1.1 Introduksjon

Reliabilitet er en utrolig viktig faktor innenfor fysiologisk testing. Hvis man ønsker å følge utviklingen til en utøver over en lenger periode er det viktig at testen vi benytter oss av, og utstyret som brukes i testen måler tilnærmet likt hver gang. Hvis testene som blir brukt har høy reliabilitet kan utøvere og trenere stole på at forskjellene i resultater mellom ulike tester skyldes endringer i fysiologiske faktorer og ikke feilmålinger som gir utslag.

For å i ettertid kunne evaluere effekten av en treningsplan, intervensjon eller periode må man kunne stole på testene som blir gjennomført. Dersom testene som blir benyttet har lav reliabilitet, kan det være vanskelig å skille mellom virkelige prestasjonsforbedringer og tilfeldige variasjoner som skyldes unøyaktighet i målingene. Dette kan resultere i at man endrer et godt fungerende treningsopplegg, eller at man fortsetter med et dårlig fungerende treningsopplegg.

I idrettsvitenskapen forskes det gjerne på effekt av ulike intervensjoner. Uavhengig av om det gjøres forskning på utrente eller elite-utøvere er det viktig at målingene har høy reliabilitet. Dette med bakgrunn i at vi vil levere god kvalitet i forskningen og at det skal være litteratur man skal kunne stole på.

Det ble gjennomført fire testdager 28.08.2024, 29.08.2024, 9.09.2024 og 11.09.2024 for å teste  $\dot{V}O_{2maks}$ . Formålet med disse testene var å øve på å kunne gjennomføre fysiologiske tester med høy reliabilitet. Reliabilitet refererer til graden av konsistens eller pålitelighet i målinger evnen til å kunne reproducere<sup>1</sup>, et eksempel på dette er ved fysiologisk testing som repeteres i forskningsprosjekter, der bedre reliabilitet vil indikere hvor god presisjonen er og måling av endring over tid<sup>1</sup>. Det er mange begreper som er relevante for å kunne si noe om reliabilitet, men standardavviket er et av disse. Standardavviket sier noe om hvor langt unna verdiens gjennomsnittlige avstand er fra gjennomsnittet<sup>2</sup>.

Kroppens maksimale oksygenopptak ( $\dot{V}O_{2maks}$ ) sier noe om kroppens maksimale evne til å ta opp og omsette oksygen<sup>3</sup>.  $\dot{V}O_{2maks}$  kan beskrives ved hjelp av Ficks likning:  $\dot{V}O_{2maks} = MV_{maks} \times a-vO_2$  differansemaks.  $\dot{V}O_{2maks}$  måles ved at man måler hvor mye oksygen kroppen klarer å omsette pr minutt<sup>3</sup>. Det finnes ulike måter og fremstille  $\dot{V}O_{2maks}$  på, de to av disse er absolutt  $\dot{V}O_{2maks}$  beskrevet som (ml  $\times$  min<sup>-1</sup>) eller relative tall relatert til kroppsvekt (ml/kg/min).

I resultatdelen har vi valgt å bruke relativ  $\dot{V}O_{2maks}$  for å beregne reliabiliteten til testene vi har gjennomført. Vi har også valgt å se på sammenhengen mellom relativ  $\dot{V}O_{2maks}$  og wattmaks under  $\dot{V}O_{2maks}$ -testen. Forskning viser at høy  $\dot{V}O_{2maks}$ , sammen med god mekanisk effektivitet og høy laktattærskel gir bedre utholdenhetsprestasjoner, noe som reflekteres i høyere  $W_{maks}/kg$ <sup>4</sup>.

## 1.2 Metode

$\dot{V}O_{2maks}$ -testen ble gjennomført på en ergometersykel med bukkestyre (Lode Excalibur Sport; Lode B.V., Groningen, Nederland). Kranken kalibreres på Lode-sykkelen før hver teststart. Dette gjøres for å få nøyaktige tråkkdata på hver forsøksperson. Sykkel stilles inn etter utøvers ønske for å sikre best mulig sittestilling ved første test. Sykkel stilles inn etter nøyaktig samme mål ved senere tester for å gjøre reliabiliteten høy. For å måle det maksimale oksygenopptaket ble det brukt Vyntus (Jaeger Vyntus CPX, Hoechberg, Tyskland). Gassanalysator kalibreres til  $< 2,0\%$  differanse og luftvolum kalibreres til  $< 0,2\%$  differanse. Syklistene veies med de klærne de skal sykle med, og 0,3kg trekkes fra (300g er et estimat på vekten av klærne forsøkspersonen har på). For å kunne sikre god relabilitet ble det tydeliggjort at man skulle replisere det siste måltidet før test, ha det samme koffeininntaket, avstå fra alkohol og tobakk de siste 72 timene før test og prøve å få tilnærmet lik søvn, samt trene det samme dagen før test. Da dette er faktorer som kan spille inn på prestasjon og metabolismen<sup>5</sup> og dermed påvirke relabiliteten.

$\dot{V}O_{2maks}$ -testen gjennomføres etter en fem minutters standardisert oppvarming på ergometersykkelen. Oppvarmingen starter to minutter på 11-12 i Borg, deretter to minutter på 15 i Borg før ett minutt på 11-12 Borg. Testen starter på en belastning (Watt) basert på deltagerens nivå i samråd med utøver og testleder. Det viktigste er at påfølgende  $\dot{V}O_{2maks}$  tester starter på samme watt. Wattbelastningen økte med 20W eller 25W (20W for kvinne og 25W for mann) hvert minutt frem til utøveren når maksimal utmattelse. Maksimal utmattelse ble i denne sammenheng ikke evne til å kunne opprettholde  $RPM > 60$ . Under  $\dot{V}O_{2maks}$  var  $RPM$  valgfritt. Testleder gjør verbal oppmuntring og sekundering underveis i testen. For at verbal oppmuntring og instruksjon ved test skulle være lik etterstrebt vi å ha samme testleder til samme forsøksperson<sup>6</sup>. Det blir registrert nye oksygenmålinger hvert 30 sek, og de to høyeste påfølgende målingene blir definert som  $\dot{V}O_{2maks}$ . Umiddelbart etter test oppgir utøveren opplevd anstrengelse på Borg skala. Maks hjertefrekvens blir lest av fra utøvers pulsklokke. Blodprøve ble tatt fra utøverens fingertupp 1 min etter endt test for å måle [BLa-]. [BLa-] blir deretter målt ved hjelp av en Biosen C-line (Biosen C-line Lactate Analyzer, EKF Diagnostic GmbH, Barleben, Germany). Etter endt test ble det hentet ut data som videre ble plottet inn i Excel og videre ført statistikk på ved hjelp av Rstudio.

## 1.3 Resultat

Etter at testene er gjennomført kan vi se nærmere på hver forsøksperson sine data, ved T1. T1 er valgt da denne testen hadde flest deltakere. Dette gir muligheten til å se på hver enkelt forsøksperson om man finner dette interessant (Table 1.1). Verdiene man kan se i tabellen, er verdier som er plottet etter endt  $\dot{V}O_{2maks}$ -test.

Etter å ha gjennomført  $VO_{2maks}$ -testene ser vi at kvinnene på 1MAIDR har et gjennomsnittlig oksygenopptak på  $3163 \pm 484$ . Mennene har derimot et gjennomsnittlig oksygenopptak på  $4380 \pm 515$ .

Reliabiliteten mellom t1 og t2 er 2.47%, mens reliabiliteten mellom t3 og t4 er 4.78%.

Table 1.1

[H]																
Parameter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Borg <sub>max</sub>	19.2 (0.96)	19 (0.82)	18 (1.2)	19 (0)	19.5 (0.71)	19 (0)	17.5 (1.7)	17 (NA)	19.7 (0.58)	20 (0)	17.5 (0.71)	18 (1.7)	18.3 (0.58)	18.8 (0.5)	17 (1)	19.5 (0.71)
VO <sub>2max</sub> (ml/kg/min)	33.5 (1.5)	43.7 (2.6)	51.6 (4.1)	37.1 (1.1)	58.9 (0.64)	45.5 (0.2)	61.8 (1.9)	43.5 (NA)	58.8 (0.59)	43.2 (0.89)	56.5 (0.94)	61.7 (3.1)	51.3 (0.88)	65.7 (1.1)	39.8 (2.6)	60.2 (1.2)
Watt/kg	2.5 (0.14)	3.58 (0.044)	3.6 (0.46)	3 (0.2)	5.18 (0.082)	3.51 (0.1)	5.24 (0.2)	3.93 (NA)	4.92 (0.038)	3.76 (0.014)	4.93 (0.049)	5.6 (0.4)	3.87 (0.062)	5.51 (0.1)	2.85 (0.12)	4.63 (0.065)
VO <sub>2max</sub> (ml/min)	3240 (150)	2700 (160)	4130 (300)	2860 (52)	4390 (48)	3710 (6.4)	5130 (140)	2540 (NA)	4650 (41)	3100 (64)	3640 (97)	4480 (230)	4590 (48)	4520 (59)	4100 (270)	4960 (130)
Watt <sub>max</sub>	243 (13)	221 (2.8)	288 (36)	231 (13)	387 (6.1)	286 (7.5)	435 (19)	230 (NA)	389 (2.1)	269 (1.2)	318 (0)	407 (29)	347 (6.9)	380 (5.7)	293 (12)	382 (3.1)

Table 1.1: Tabellen viser hver deltakers gjennomsnitt og standardavvik i () på verdier vi har undersøkt

### 1.3.1 Korrelasjon mellom $\dot{V}O_{2maks}$ og Wattmaks per kg

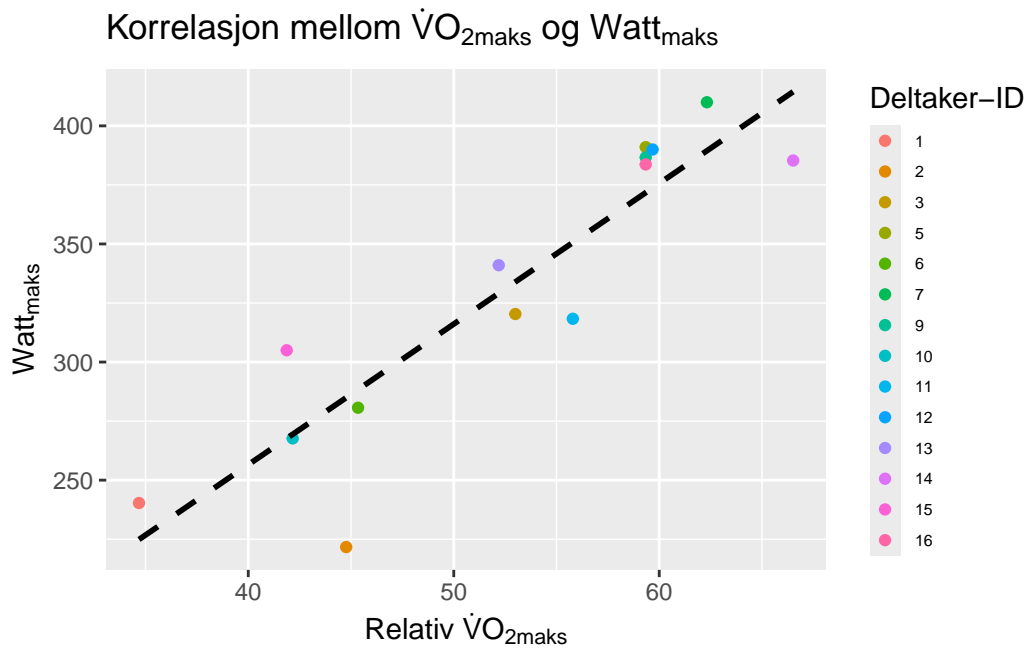


Figure 1.1: Hvert punkt = én observasjon

## 1.4 Diskusjon

På bakgrunn av resultatene vi har observert under testing av  $\dot{V}O_{2maks}$ , kan man slå fast at reliabiliteten til metoden vi har brukt er ganske god. Resultatene tyder på at vi får målt det vi er ute etter på en god måte, og det vil variere lite i måle-utstyret fra gang til gang.

Dette sikrer at vi med ganske god sikkerhet, kan fastslå om et treningsprogram fungerer, ved å gjøre repeterte tester, med treningsperioder i mellom.

## 2 Regresjonsmodeller og prediksjoner

### 2.1 Introduksjon

En regresjonsmodell er en modell som kvantifiserer forholdet mellom en eller flere uavhengige variabler og en avhengig variabel. Innen medisin er regresjon den analysemetoden som er hyppigst anvendt. Det finnes forskjellige regresjonsmodeller. De vanligste er lineær regresjon, polynominal regresjon og logistisk regresjon. Hva man har av datasett vil bestemme hvilken regresjonsmodell som egner seg best å benytte<sup>7</sup>.

En lineær regresjonsmodell er en modell der en kan estimere verdien av en avhengig variabel basert på verdien av andre kjente uavhengige variabler<sup>7</sup>. I en slik modell benyttes en rett linje for å lage en modell som beskriver dataen. Følgende funksjon benyttes for å skape det lineære plottet:

$$y_i = b_0 + b_1 x_i + e_i$$

der  $y_i$  er den avhengige variabelen som kan estimeres ved å benytte de uavhengige variablene  $b_1 x_i$  og  $b_0$ .  $b_0$  er skjæringspunktet til grafen og  $b_1$  er stigningstallet til grafen.

Laktatterskeltesting er en sentral metode innen prestasjonsanalyse og treningsoptimalisering, spesielt i utholdenhetsidretter. Laktatterskelen representerer den høyeste intensiteten hvor kroppen kan opprettholde balanse mellom produksjon og eliminering av laktat i blodet. Dette er en viktig parameter for å forstå hvordan kroppen responderer på økende treningsintensitet, og den gir verdifull innsikt i både fysisk kapasitet og treningsprogrammering.

For å bestemme laktatterskelen benyttes typisk en trinnvis økning i treningsintensitet mens blodprøver tas for å måle laktatnivåer. Relasjonen mellom treningsintensitet og blodlaktat analyseres deretter ved hjelp av regresjonsmodeller for å estimere intensitetsnivåene ved faste laktatkonsentrasjoner, ofte 2 og 4 mmol/L. Disse tersklene gir grunnlag for å skille mellom moderat og høy intensitet i trening, noe som er avgjørende for å optimalisere treningsplaner<sup>8</sup>.

### 2.2 Part 1 - Lactate thresholds

#### 2.2.1 Metode

Dataene ble organisert i et mer hensiktsmessig format (tidy data) for å forenkle videre analyse og modellering. Deretter ble ulike regresjonsmodeller anvendt for å representere dataene. Nye skjæringspunkter ble tegnet opp for å illustrere treningsintensitet ved forskjellige laktatnivåer.



## 2.2.2 Resultat

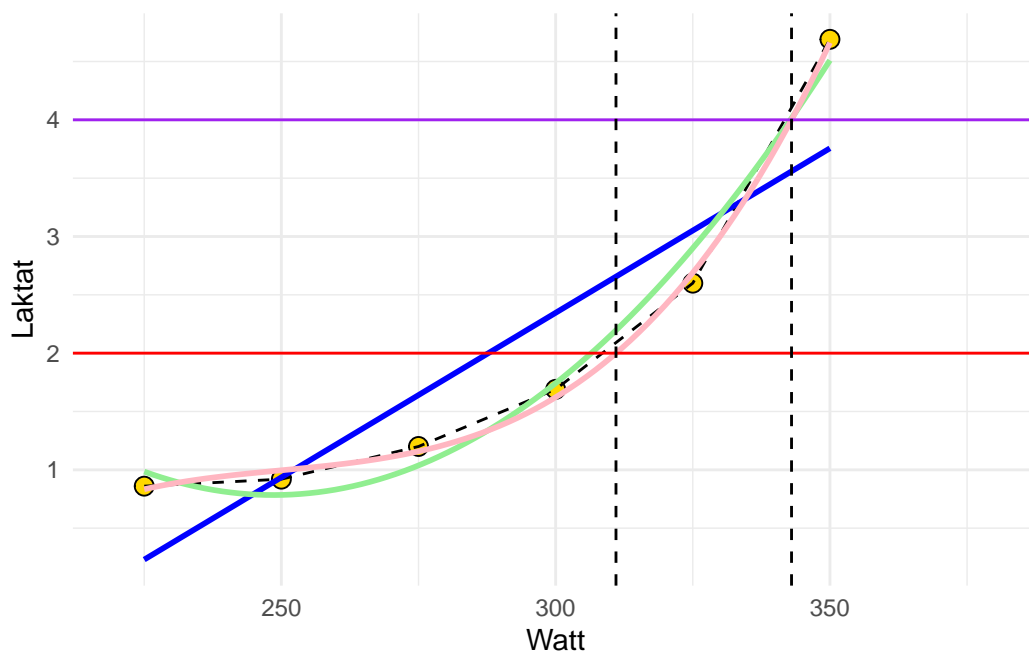


Figure 2.1: Figur 1: Gule punkter: laktat og watt, blå linje: lineær regresjon, lysegrønn: andregradsligning, rosa: tredjegradsligning.

Vår beste prediksjon til belastningen på  $2 \text{ mmol} \times \text{L}^{-1}$  er 311W, denne kommer fra tredjegradsligningen. Den predikerte belastningen fra en lineær modell gir 288W. For predikert terskelwatt ( $4 \text{ mmol} \times \text{L}^{-1}$ ) gir tredjegradsligningen 343W, som er den beste prediksjonen, mens den lineære modellen gir et estimat på 359W.

## 2.2.3 Diskusjon

Vi har valgt å se på subject 10 fra datasettet Cyclingstudy. Vi gjør om datasettet til tidydata. Dette gjør vi for å gi watt og laktat hver sine verdier. Vi plotter inn laktatverdier og wattverdier (gule punkter). Deretter tegner vi en stiplet linje som følger punktene. Vi gjør en regresjonsanalyse, først en lineær modell (blå linje), deretter en andregradsligning (grønn) og til slutt en tredjegradsligning (rosa). Disse bruker vi for å observere hvilken modell som passer best i dette tilfellet.

For å understreke hvor unøyaktig den lineære modellen er i dette tilfellet, kan man på øyemål se at laktaten på 300W viser omtrent  $2.4 \text{ mmol} \times \text{L}^{-1}$ . Den faktiske laktaten på 300W er  $1.69 \text{ mmol} \times \text{L}^{-1}$  (Figure 2.1).

## 2.3 Part 2 - Predicting sizes of DNA fragments

### 2.3.1 Metode

For å predikere kalibreringskurven til qPCR, må en rekke prosesser på molekylærlaboratoriet gjennomføres før dataene kan analyseres i R Studio.

For å utføre en PCR på en 2% agarosegel, ble det først tatt helblod fra en forsøksperson for å ekstrahere DNA. Helblodet gjennomgikk ulike prosesser hvor forskjellige løsninger og primere ble tilsatt. Dette resulterte i et PCR-produkt. En elektroforese ble deretter kjørt for å separere DNA-fragmentene fra PCR-reaksjonen. Etter fullført elektroforese ble det tatt et bilde av 2% agarosegelelen.

Bildet fra elektroforesen ble analysert ved hjelp av ImageJ/Fiji, og videre dataanalyser ble utført i R og R Studio. PCR-reaksjoners effektivitet bestemmes av primerdesign og deres spesifisitet.

### 2.3.2 Resultat

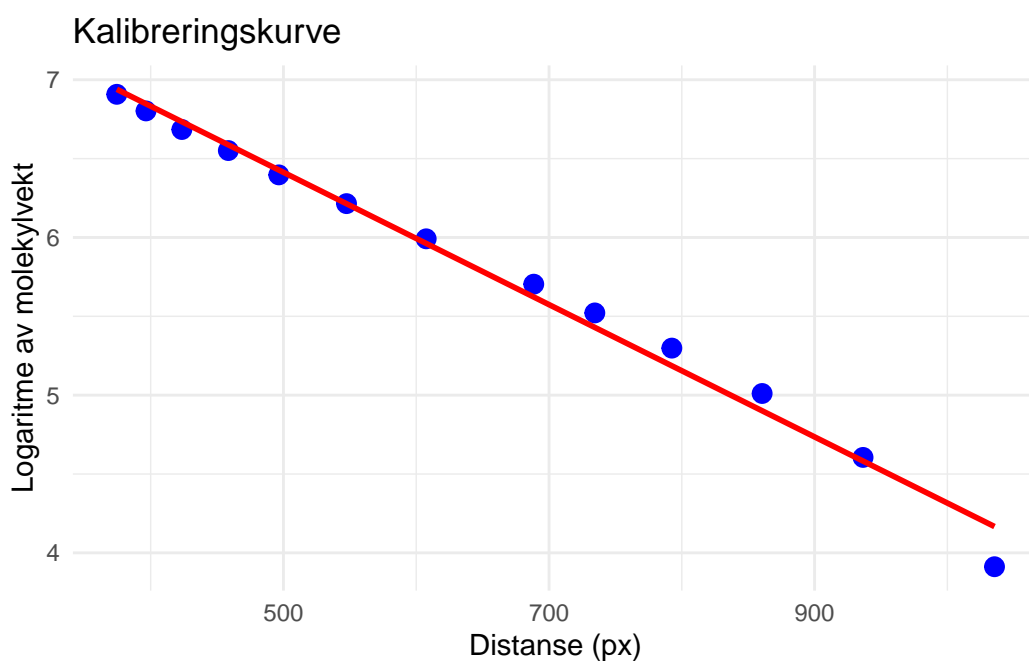


Figure 2.2

Table 2.1: Predikerte molekylvekter for ukjente distanser

Båndnummer	Distanse (px)	Predikert molekylvekt (bp)
1	1208.5	31.22
2	600.5	400.05
3	18.5	4595.75
4	383.5	994.09
5	408.5	895.12
6	436.5	795.93
7	470.5	690.14
8	508.5	588.45
9	559.5	475.12
10	618.5	370.95
11	696.5	267.44
12	742.5	220.51
13	798.5	174.34
14	862.5	133.3
15	935.5	98.14
16	993.5	76.94

### 2.3.3 Diskusjon

I denne delen av oppgaven har vi benyttet en kalibreringskurve for å predikere størrelsene på ukjente DNA-fragmenter fra en agarosegelelektroforese. Ved å bruke et kjent DNA-ladder med kjente molekylvekter, etablerte vi en lineær sammenheng mellom logaritmen av molekylvekten og migrasjonsavstanden på gelen.

Resultatene viser en tydelig lineær korrelasjon mellom  $\log(mw)$  og distanse, noe som er i tråd med teori, da mindre DNA-fragmenter migrerer lenger gjennom gelen enn større fragmenter. Den tilpassede lineære modellen ble deretter brukt til å estimere molekylvektene til de ukjente fragmentene basert på deres migrasjonsavstander (Figure 2.2).

Tabellen presenterer de predikerte molekylvektene for hvert ukjent bånd, og fargekodingen gir en visuell indikasjon på fragmentenes relative størrelser (Table 2.1). Dette er nyttig for raskt å identifisere fragmenter av interesse, spesielt i komplekse prøver med mange bånd.

Det er viktig å være oppmerksom på at nøyaktigheten av prediksjonene avhenger av flere faktorer. Kvaliteten på gelen, nøyaktigheten i måling av migrasjonsavstander, og lineariteten i området av molekylvekter som analyseres, kan alle påvirke resultatene. For eksempel kan avvik i gelens konsistens eller løpeforhold føre til uregelmessig migrasjon av DNA-fragmenter, noe som kan gi feilaktige estimater.

Videre antar den lineære modellen en eksponentiell sammenheng mellom migrasjonsavstand og molekylvekt, noe som kan være mindre nøyaktig for svært store eller små fragmenter. I slike tilfeller kan det være fordelaktig å bruke en ikke-lineær modell eller å inkludere flere kalibreringspunkter for å forbedre presisjonen.

Samlet sett demonstrerer denne analysen effektiviteten av å bruke gelelektroforese kombinert med kalibreringskurver for å estimere størrelsen på ukjente DNA-fragmenter. Dette er en grunnleggende

teknikk i molekylærbiologi som er essensiell for en rekke applikasjoner, inkludert genotyping, kloning og diagnostikk.

## 2.4 Part 3 - Interpreting a regression table

### 2.4.1 Metode

I denne analysen undersøkte vi sammenhengen mellom treningsalder (antall år med trening) og testosteronkonsentrasjon i blodet ved tidspunkt T1 (målt i  $\text{ng} \times \text{dl}^{-1}$ ). Vi benyttet datasettet *hyper-trophy* fra *exscidata*-pakken i R. Variablene som ble analysert var `TRAINING_AGE` (treningsalder) og `TESTOSTERONE_T1` (testosteronnivå ved T1).

En enkel lineær regresjonsanalyse ble utført med testosteronnivå som avhengig variabel og treningsalder som uavhengig variabel. Regresjonsmodellen ble tilpasset ved hjelp av funksjonen `lm()` i R:

```
m <- lm(TESTOSTERONE_T1 ~ TRAINING_AGE, data = dat)
```

Vi hentet ut regresjonskoeffisientene, standardfeilen, t-verdien og p-verdien fra modellens sammendrag ved hjelp av `summary()`-funksjonen. Vi estimerte også det forventede testosteronnivået etter 10 år med trening ved å bruke regresjonsligningen.

### 2.4.2 Resultat

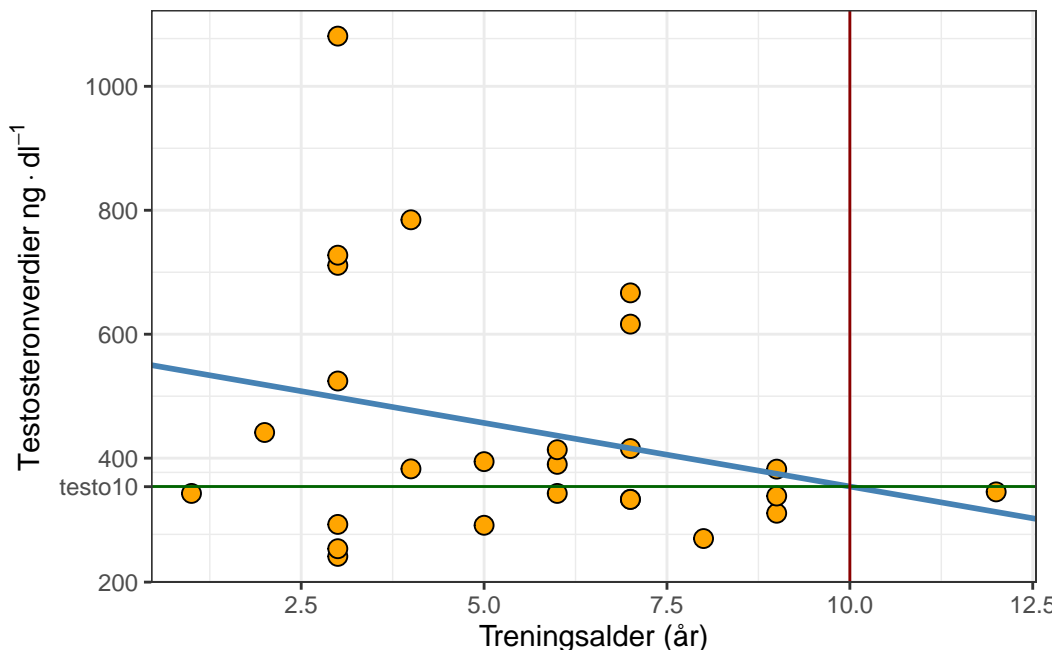


Figure 2.3: Figur 3: Sammenheng mellom treningsalder og testosteronverdier i blodet

Den lineære regresjonsmodellen viste at det var en negativ sammenheng mellom treningsalder og testosteronkonsentrasjon i blodet (Figure 2.3). Regresjonskoeffisienten for treningsalder var  $-20.51 \text{ ng} \times \text{dl}^{-1}$  per år, noe som indikerer at testosteronnivået synker med gjennomsnittlig  $20.51 \text{ ng}$

$\times \text{dl-1}$  for hvert år med trening. Interceptet i modellen var  $559.36 \text{ ng} \times \text{dl-1}$ , som representerer det estimerte testosteronnivået for en person uten treningserfaring.

T-verdien for koeffisienten var  $-1.39$ , og tilhørende p-verdi var  $0.178$ . Siden p-verdien er høyere enn signifikansnivået på  $0,05$ , er sammenhengen ikke statistisk signifikant. Dette betyr at vi ikke har tilstrekkelig bevis til å konkludere med at det er en lineær sammenheng mellom treningsalder og testosteronnivå i denne populasjonen.

Ved å bruke regresjonsligningen estimerte vi testosteronnivået etter 10 år med trening til å være  $354.26 \text{ ng} \times \text{dl-1}$ . Dette ble beregnet ved å sette inn  $\text{TRAINING\_AGE} = 10$  i modellen.

Dette estimatet illustrerer den forventede nedgangen i testosteronnivå basert på modellen, men gitt den ikke-signifikante p-verdien bør resultatene tolkes med forsiktighet.

### 2.4.3 Diskusjon

Fra datasettet hypertrophy valgte vi å se på sammenhengen mellom testosteronkonsentrasjon i blodet ( $\text{ng} \times \text{dl-1}$ ) og treningsalder (antall år med trening). Den lineære modellen forteller at testosteronkonsentrasjonen i blodet synker med  $20.51 \text{ ng} \times \text{dl-1}$  for hvert treningsår. Etter 10 år med trening, estimerer den lineære modellen et testosteronnivå på  $354.26 \text{ ng} \times \text{dl-1}$ .

Analysen av dataene viser en p-verdi på  $0.178$ , noe som indikerer at det ikke er statistisk signifikant bevis for en sammenheng mellom treningsalder og nivået av testosteron i blodet. Siden p-verdien er høyere enn det vanlige signifikansnivået på  $0,05$ , kan vi ikke avvise nullhypotesen, som antyder at det ikke er noen betydelig effekt eller sammenheng mellom de to variablene i dette datasettet. Dette betyr at variasjonen i testosteronnivåer ikke ser ut til å være relatert til hvor lenge individene har trent.

I analysen av sammenhengen mellom treningsalder og testosteronnivåer i blodet ses det en t-verdi på  $-1.39$ . Den høye t-verdien på  $-1.39$ , og en p-verdi på  $0.178$ . Denne p-verdien er høyere enn det vanlige signifikansnivået på  $0,05$ , noe som betyr at vi ikke har tilstrekkelig statistisk bevis for å avvise nullhypotesen. Selv om t-verdien indikerer en mulig sammenheng mellom treningsalder og testosteronnivå, er det ikke nok evidens til å konkludere med at denne sammenhengen er signifikant. Dermed kan vi konkludere med at selv om det kan være en tendens til en sammenheng mellom treningsalder og testosteronnivåer, er resultatene fra denne analysen ikke sterke nok til å si at treningsalder har en reell effekt på testosteronnivåene i blodet.

## 3 Å trekke slutninger fra statistiske modeller og statistisk styrke

### 3.1 Spørsmål og svar

#### 3.1.1 Estimate

Et **estimat** er en verdi vi får ved å anvende en lineær modell på våre data. Dette tallet representerer vår beste gjetning av den sanne verdien til parameteren vi ønsker å estimere i populasjonen, basert på vårt utvalg. I konteksten av en regresjonsanalyse er dette ofte regresjonskoeffisienten, som estimerer sammenhengen mellom en uavhengig variabel og den avhengige variabelen  $y$ .

**Standardfeilen** (SE) kvantifiserer usikkerheten knyttet til estimatet vårt. Den måler den forventede variasjonen i estimatet dersom vi skulle trekke mange utvalg fra populasjonen og beregne estimatet hver gang. Med andre ord er standardfeilen standardavviket til estimatets utvalgsfordeling, og indikerer hvor mye estimatet vårt potensielt kan variere fra utvalg til utvalg på grunn av tilfeldig sampling.

**t-verdien** er forholdet mellom estimatet og standardfeilen ( $t = \frac{\text{estimat}}{SE}$ ). Den indikerer hvor mange standardfeil estimatet er unna null<sup>2</sup>. En høy absolutt t-verdi tyder på at estimatet er signifikant forskjellig fra null.

**P-verdien** angir sannsynligheten for å observere en t-verdi som er minst like ekstrem som den vi har fått, gitt at nullhypotesen er sann. Det vil si, den måler sannsynligheten for å få våre data, eller data som er mer ekstreme, dersom det faktisk ikke er noen effekt (dvs. hvis den sanne parameteren er null). En lav p-verdi indikerer at et så ekstremt resultat er lite sannsynlig under nullhypotesen, noe som gir grunnlag for å forkaste nullhypotesen<sup>2</sup>.

I vårt tilfelle har vi en høy p-verdi, noe som indikerer at vi ikke kan forkaste nullhypotesen. Dette betyr at det ikke er tilstrekkelig bevis til å konkludere med at det er en signifikant forskjell fra null.

#### 3.1.2 $m1$ vs $m2$

Forskjellen mellom studiene kommer fra størrelsen på utvalget som er brukt i de to forskjellige. I  $m1$  er det brukt ett mye mindre utvalg, noe som fører til større usikkerhet rundt resultatene. I  $m2$  er det brukt et større utvalg, som gjør at det estimerte gjennomsnittet blir nærmere populasjonsgjennomsnittet og standardfeilen blir dermed mindre. Dette gir i vårt tilfelle en høyere **t-verdi** og en lavere **p-verdi**<sup>2</sup>.

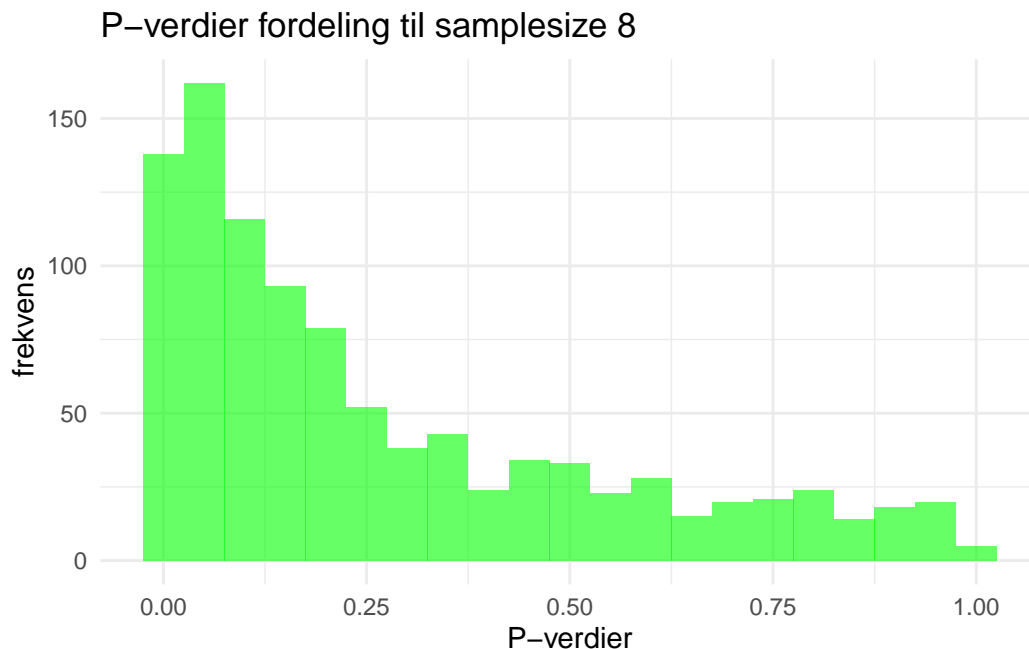
### 3.1.3 Shaded areas

Vi bruker de grå feltene for å vise de ekstreme verdiene vi har fra testen vår. Jo lenger ut i halene vi kommer, desto større sannsynlighet er det for at dette er et uvanlig resultat å se.

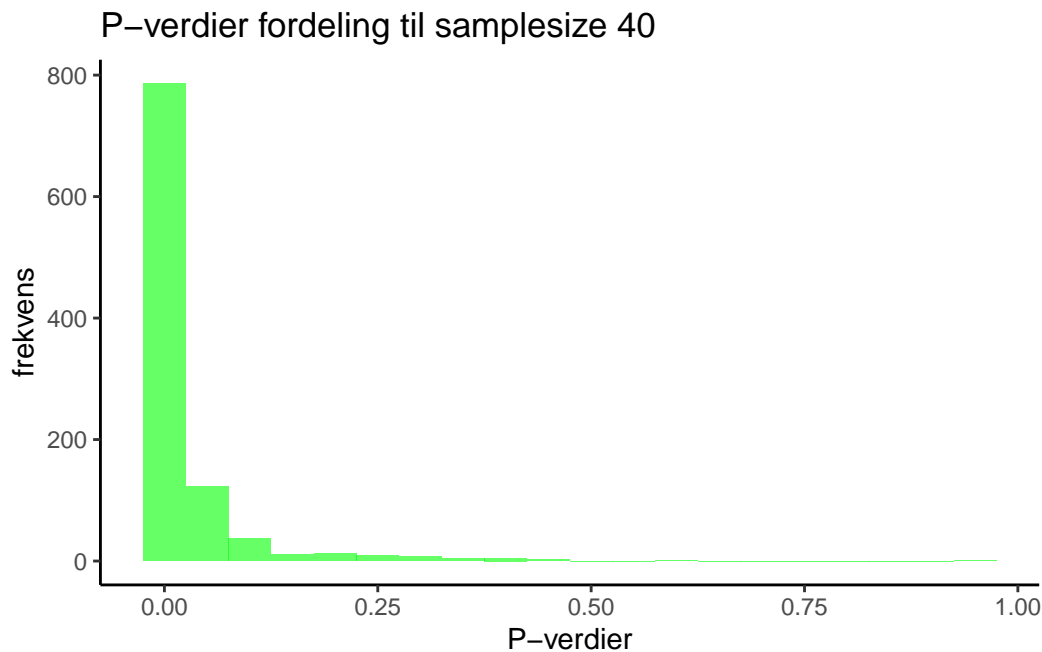
### 3.1.4 Standard deviation of estimate and avg. se for each study.

Standard deviation for modellen med 8 i population er 1.07, mens det for modellen med 40 i population er 0.48. Når det kommer til gjennomsnittlig standardfeil ligger den på 1.02 for modellen med 8 i population, mens den for modellen med 40 i population ligger på 0.47. Grunnen til at tallene er såpass like som de er for **SD** og **avg se** er at begge beregningene er mål på variasjon. I denne sammenhengen er standardfeilen et mål på hvor mye gjennomsnittet avviker fra det sanne populasjonsgjennomsnittet.

### 3.1.5 P-value histogram



Når vi ser histogrammet for modellen med utvalgsstørrelse på 8, ser vi tydelig at det er mange observasjoner av høye p-verdier. Dette gjenspeiler den lave statistiske poweren vi får av å gjøre studier med en så liten utvalgsstørrelse.



På histogrammet med utvalgsstørrelse på 40 ser vi at det er en mye større samling av observasjoner på lave p-verdier. Dette gjenspeiler det vi vet om at en større utvalgsstørrelse gir en større statistisk power.

### 3.1.6 Antall studier med statistisk signifikans

I studiene med utvalgsstørrelse på 8 ser vi at det er 227 studier som viser statistisk signifikans, mens det i studiene med utvalgsstørrelse på 40 er hele 865 studier som viser statistisk signifikans. Dette gir et godt bilde på hvor mye utvalgsstørrelsen har å si for resultatet i utregningen vår. I mitt tilfelle har jeg valgt å sette terskelen for signifikans (p-verdi) til 0.05.

### 3.1.7 Power of a one-sample t-test

Når vi gjennomfører utregningen ser vi at studiene med lav utvalgsstørrelse (8) får en mye lavere statistisk styrke (0.232) enn studiene med utvalgsstørrelse på 40 (0.869). Svarene vi får av disse utregningene støtter det vi tidligere har funnet ut, at dersom vi har et større utvalg, er det større sannsynlighet for at vi ser en faktisk effekt, og at det ikke er en tilfeldighet at vi har funnet det vi har i studien. I dette tilfelle vil vi da få en 86.9% sjanse for å oppdage en sann effekt.

### 3.1.8 Med signifikansnivå på 0.05 hvor mange studier gir “falsk positiv” ved gjennomføring av mange repeterte studier?

Ved å gjøre 1000 repeterte studier, vil vi få omtrent 50 falske positive hvis vi setter signifikansnivået til 0.05. I min utregning fikk jeg da 49 for studiene med utvalgsstørrelse på 8 og 59 på studiene med utvalgsstørrelse 40. Om jeg endrer signifikansnivået og setter alpha enda lavere vil resultatet endre seg litt. Med en signifikansverdi på 0.025 vil det i studiene med utvalgsstørrelse 8 gi meg 30 falske positive, mens det på studiene med 40 i utvalgsstørrelse gir 22 falske positive.



## 3.2 Code-chunks fra topp til bunn

```
set.seed(1)
population <- rnorm(1000000, mean = 1.5, sd = 3)

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

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

# Skjul summary-output
invisible(summary(m1))
invisible(summary(m2))
```

```
# 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 ne
# 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)

# Calculate standard deviation of the estimate and the average of the stand
results_summary <- results |>
  group_by(n) |>
  summarise(
    sd_estimate = sd(estimate),
    avg_se = mean(se)
  )

sd_est_8 <- sd(results$estimate[results$n == 8])
avg_se_8 <- mean(results$se[results$n == 8])

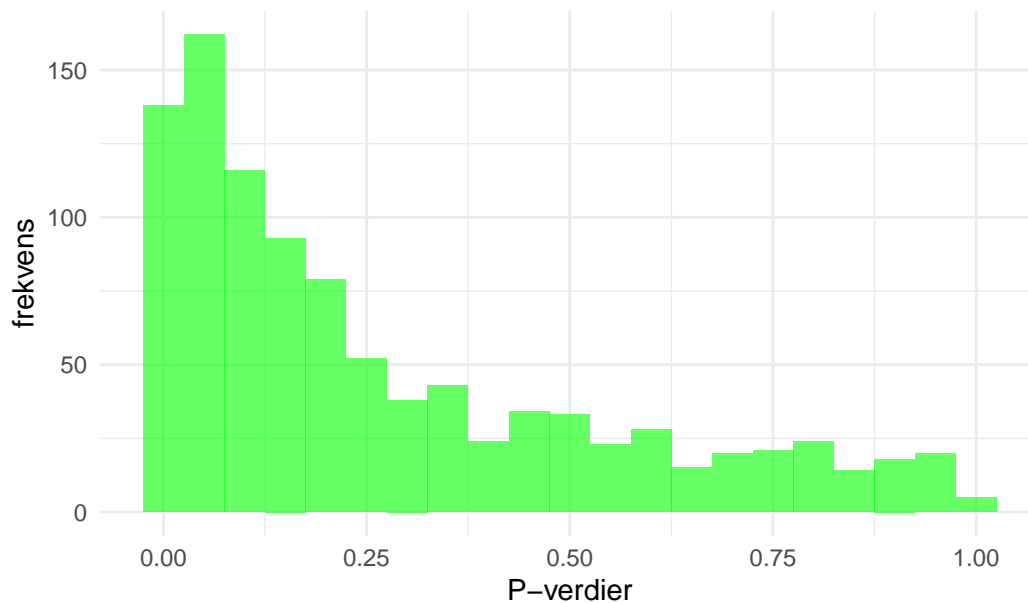
sd_est_40 <- sd(results$estimate[results$n == 40])
avg_se_40 <- mean(results$se[results$n == 40])

rounded_sd_est_8 <- round(sd_est_8, 2)
rounded_avg_se_8 <- round(avg_se_8, 2)
rounded_sd_est_40 <- round(sd_est_40, 2)
rounded_avg_se_40 <- round(avg_se_40, 2)

ggplot(results[results$n == 8, ], aes(x = pval)) +
  geom_histogram(binwidth = 0.05, fill = "green", alpha = 0.6) +
  labs(title = "P-verdier fordeling til samplesize 8",
       x = "P-verdier",
       y = "frekvens") +
  theme_minimal()

```

### P-verdier fordeling til sample size 8



```
alpha <- 0.05

significant_8 <- sum(results$pval[results$n == 8] < alpha)
significant_40 <- sum(results$pval[results$n == 40] < alpha)
```

```
effect_size <- 1.5 / 3

power_8 <- pwr.t.test(n = 8,
                     d = effect_size,
                     sig.level = alpha,
                     type = "one.sample")$power
rounded_power_8 <- round(power_8, 3)

power_40 <- pwr.t.test(n = 40,
                      d = effect_size,
                      sig.level = alpha,
                      type = "one.sample")$power

rounded_power_40 <- round(power_40, 3)

rounded_power_40_perc <- rounded_power_40 * 100
```

```
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 ne
# 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_8 <- sum(results_8$pval < 0.05)
false_positive_40 <- sum(results_40$pval < 0.05)

false_positive_8_alpha0.025 <- sum(results_8$pval < 0.025)
false_positive_40_alpha0.025 <- sum(results_40$pval < 0.025)

```

## 4 Studiedesign

### 4.1 Introduksjon

Formålet med denne sammenligningen er å undersøke ulikheter i de metodiske valgene blant ulike studier som ser på effekten av blokkperiodisering av utholdenhetstrening sammenlignet med tradisjonell utholdenhetstrening på maksimal oksygenopptak ( $\dot{V}O_{2maks}$ ). Videre vil jeg vurdere styrker og svakheter ved studiedesign, statistiske analyser og resultater. Til slutt vil jeg gi noen anbefalinger for fremtidige studier.

### 4.2 Metode

Fem ulike studier ble valgt ut, alle med fokus på effekten av blokkperiodisering av utholdenhetstrening på  $\dot{V}O_{2maks}$ . Studienes styrker og svakheter ble analysert, med spesiell vekt på studiedesign og statistiske analyser. Studiedesignenes evne til å måle relevante utfall ble vurdert, samt i hvilken grad de tok hensyn til eksterne variabler som kunne påvirke resultatene.

### 4.3 Resultat

Den første studien<sup>9</sup> brukte et kontrollert design med objektive målinger som  $\dot{V}O_{2maks}$ . Til tross for dette var utvalgsstørrelsen liten, noe som kan redusere den statistiske kraften og øke risikoen for type II-feil. Manglende kontroll for eksterne faktorer, som kosthold og annen fysisk aktivitet utenom studien, kan ha påvirket resultatene og dermed redusert intern validitet.

Den andre studien<sup>10</sup> hadde et robust eksperimentelt design med godt definerte kontrollgrupper og brukte gjentatte målinger for å spore prestasjonsendringer over tid. Likevel var ekstern validitet en utfordring på grunn av spesifisiteten til utvalget, som besto av eliteutøvere. Dette gjør det vanskelig å generalisere funnene til en bredere befolkning. I tillegg var tidsrammen for studien relativt kort, noe som begrenser muligheten til å vurdere langsiktige effekter av treningsintervensjonen.

Studie nummer tre<sup>11</sup> implementerte blokkperiodisering og benyttet et bredt spekter av prestasjonsmål. Den korte intervensjonsperioden og det lave antallet kvinnelige deltakere utgjorde imidlertid svakheter som kan ha påvirket representativiteten og dermed generaliserbarheten av konklusjonene. Dette kan også ha ført til skjevheter i resultatene grunnet kjønnsespesifikke responser på treningen.

Den fjerde studien<sup>12</sup> benyttet avanserte statistiske analyser, inkludert lineære mixed models, og hadde et langsiktig design. Likevel var utvalgsstørrelsen begrenset, og fokus på eliteutøvere reduserte generaliserbarheten av funnene til andre populasjoner, slik som mosjonister eller utøvere på lavere nivåer.

Studie fem<sup>13</sup> brukte et omfattende design med både fysiologiske og prestasjonsbaserte utfallsmål. Svakheten var spesifisiteten til deltakerutvalget, som besto av kvinnelige eliteutøvere i langrenn. Dette begrenser overførbarheten av resultatene til mannlige utøvere eller til andre idretter. Det var også behov for en bedre sammenheng mellom de statistiske testene og forskningsspørsmålene i studien; for eksempel kunne valg av mer avanserte statistiske metoder ha gitt dypere innsikt i dataene.

## 4.4 Diskusjon

Studiene har generelt god intern validitet på grunn av robuste studiedesign og bruk av kontrollgrupper, noe som styrker troverdigheten til de kausale sammenhengene som blir funnet. Intern validitet refererer til i hvilken grad studiens resultater faktisk skyldes intervensjonen og ikke andre faktorer.

Den eksterne validiteten er imidlertid begrenset. Ekstern validitet handler om i hvilken grad funnene kan generaliseres til andre populasjoner og kontekster. Små utvalgsstørrelser øker usikkerheten i estimatene og reduserer statistisk kraft, noe som kan gjøre det vanskelig å oppdage reelle effekter. Bruk av spesifikke grupper, som eliteutøvere, betyr at resultatene kanskje ikke er direkte overførbare til den generelle befolkningen eller til utøvere på lavere nivåer. Dette skyldes at eliteutøvere kan ha unike fysiologiske egenskaper og treningsresponsen.

Variasjon i varigheten av treningsintervensjonene mellom studiene gjør det utfordrende å sammenligne resultatene og trekke konklusjoner om langsiktige effekter. Kortvarige intervensjoner kan mangle tilstrekkelig tid for å observere betydelige fysiologiske endringer, mens lengre studier gir bedre innsikt i vedvarende effekter av treningen.

Når det gjelder statistiske analyser, har studiene generelt sett valgt passende metoder. Flere studier har brukt gjentatt ANOVA for å analysere endringer over tid innen og mellom grupper, noe som er egnet for design med gjentatte målinger. Andre har benyttet t-tester for sammenligning mellom to grupper. Det er viktig at valg av statistiske analyser er tett koblet til forskningsspørsmålene og studiedesignet for å sikre at analysene gir meningsfulle og presise svar. For eksempel bør studier med komplekse design og flere variabler vurdere bruk av multivariate analyser eller regresjonsmodeller som kan håndtere flere konfunderende faktorer. Manglende detaljert begrunnelse for valg av statistiske tester kan svekke forståelsen av hvordan resultatene støtter forskningshypotesene.

## 4.5 Konklusjon

For å forbedre generaliserbarheten av funnene, bør fremtidige studier fokusere på å øke utvalgsstørrelsen og inkludere mer varierte populasjoner. Dette er spesielt viktig hvis målet er å overføre resultatene til den generelle befolkningen eller til utøvere på ulike nivåer. Dersom studiene primært har som mål å generalisere til toppidrettsutøvere, bør dette tydeliggjøres i målsettingen, og metodene bør tilpasses deretter.

Lengre oppfølgingstider kan bidra til å vurdere de langsiktige effektene av blokkperiodisering og gi et mer fullstendig bilde av treningsadaptasjonene. Valg av statistiske analyser bør være nøye tilpasset forskningsspørsmålene og studiedesignet. Ved å velge analyser som gjenspeiler kompleksiteten i dataene og tar hensyn til potensielle konfunderende variabler, kan studiene gi mer pålitelige og

anvendbare resultater. Det er også viktig å begrunne valgene av statistiske tester og diskutere deres begrensninger for å styrke validiteten og reliabiliteten til funnene.

## 5 Analysering av eksperimenter med gjentatte målinger

### 5.1 Introduksjon

Et styrketreningsprogram kan bestå av mange ulike variabler, som i teorien skal påvirke adaptasjoner. Volum, intensitet, frekvens, pauselengder, samt ernæring, kontraksjonstype og kontraksjonshastighet er eksempler på dette. At vi har så mange forskjellige variabler, gjør at vi har muligheten til å gjøre endringer på utallige forskjellige måter for å manipulere og tilpasse treningsprogrammer. Volum i styrketrening er et mye debattert tema, her er det spesielt ett sett, mot flere sett som har fått mye oppmerksomhet<sup>14</sup>.

Flere studier har vist at økt treningsvolum er fordelaktig for både muskelstyrke og muskelvekst (hypertrofi)<sup>15,16</sup>. Likevel finnes det forskning som indikerer at lavt volum kan gi styrke- og masseøkninger som er sammenlignbare med de som oppnås ved moderat volum<sup>17,18</sup>. Denne variasjonen i studieresultater skyldes sannsynligvis en kombinasjon av små utvalgsstørrelser og individuelle forskjeller. Studiedesign som sammenligner ulike treningsvolumer hos samme person kan teoretisk sett bidra til å håndtere disse begrensningene. I flere studier som undersøker ett sett kontra tre sett, er det også forskjeller i intensitet og hvilke øvelser som benyttes<sup>19,20</sup>.

Formålet med analysene i denne rapporten var å sammenligne effekten av ett sett versus flere sett på både muskelstyrke og hypertrofi. Med tanke på de metodiske utfordringene i studier som sammenligner ett sett med flere sett, er følgende hypotese formulert: Tre sett vil være mer effektive for å forbedre maksimal muskelstyrke og øke muskelmasse sammenlignet med ett sett.

### 5.2 Metode

#### 5.2.1 Forsøkspersoner og studiedesign

Det ble rekruttert 41 mannlige og kvinnelige deltakere, kriteriene for å bli inkludert i studien var å ikke røyke, samt være mellom 18 og 40 år. Eksklusjonskriteriene var at vedkommende hadde trent mer enn én styrkeøkt i uken, i løpet av de siste 12 månedene før intervensjonen startet, intoleranse mot bedøvelse, reduksjon i muskelstyrke grunnet skade og bruk av reseptbelagte medisiner som i verstefall kunne påvirke treningsadaptasjoner. Det var syv deltakere som ble ekskludert fra analysene som ble gjort, fordi de ikke fullførte minst 85% av den planlagte treningen. Samtlige inkluderte deltakere rapporterte at de hadde erfaring med idrettsaktiviteter fra tidligere. Blant deltakerne var det tjue av dem som drev med fysisk trening da de meldte seg på studien; blant disse var det ti av dem som drev sporadisk styrketrening, men felles for dem var at ingen trente mer enn én gang i uken.



Intervensjonen besto av 12 uker med styrketrening for hele kroppen, denne ble fullført av samtlige deltakere fra september til november. Det ble gjort en randomisering på hvert ben hos deltakerne, for å muliggjøre differensiering av treningsvolum hos samme deltaker. Hver deltaker fikk da tilfeldig tildelt enten ett eller tre sett, til hvert av beina sine, så hver person fikk fulgt begge protokollene. Det ble gjort måling av muskelstyrke ved baseline, underveis (uke 3, 5 og 9) og etter intervensjonen. Målinger av kroppssammensetning ble utført før og etter intervensjonen.

### **5.2.2 Treningsprotokoll**

Styrkeøvelsene ble gjennomført i denne rekkefølgen: unilateral beinpress, beincurl og kneekstensjon. Ett sett op det ene beinet, og tre sett på det andre beinet, ut i fra randomiseringen. Beinene som skulle trene ett sett, ble trent mellom andre og tredje sett på beinet som trente tre sett. Når øvelsene på beina var gjennomført, trente de også to sett av bilateral benkpress, nedtrekk og enten sittende roing eller skulderpress (skulderpress og sittende roing varierte fra økt til økt (annen hver gang)). Pauselengde mellom settene var mellom 1.5 og 3 minutter. Det ble gjort en gradvis økning i treningsmotstanden utover treningsintervensjonen. Deltakerne startet med 10 RM de to første ukene, deretter 8 RM i tre uker og 7 RM i syv uker. Etter økt nummer ni, ble motstanden redusert på én av de tre øktene som var ukentlig. Dette var en reduksjon tilsvarende 90 % (av motstanden) fra forrige økt på den gitte øvelsen. Deltakeren hadde fortsatt mål om samme repetisjonsantall. Det ble satt et minimum om 48 timer fra fullført økt med maksimal innsats og frem til neste økt. Etter styrkeøkene med redusert motstand var det minst 24 timer til den neste økten. For å sørge for best mulig umiddelbar resitusjon fikk deltakerne en standardisert drikk etter hver gjennomførte økt med 0.15 g/kg protein, 1.2 g/kg karbohydrater og 0.5 g/kg fett.

### **5.2.3 Målinger av muskelstyrke og hypertrofi**

Maksimal styrke er definert som den motstanden man maksimalt klarer å løfte en repetisjon av (1 RM) i beinpress og kneekstensjon. Det ble gjort en spesifikk oppvarming med ti, seks og tre repetisjoner på henholdsvis 50, 75 og 85 % av forventet 1 RM. Motstanden ble deretter økt progressivt helt til deltakeren ikke lenger klarte å løfte mer. Den høyeste motstanden med godkjent repetisjon er definert som 1 RM. Deltakerne fikk fire til seks forsøk hver.

Testene ble gjennomført to ganger ved baseline, med fire dager mellom. Den høyeste enkeltverdien de oppnådde på disse to dagene er brukt i analysene. Styrketestene ble gjort minst 48 timer etter siste gjennomførte økt etter intervensjonen. Ikke alle deltakerne ( $n = 18$ ) gjorde styrketestene underveis i intervensjonen (uke to, fem og ni). Det ble prioritert trening for de andre deltakerne, dersom de gikk glipp av enten test eller trening grunnet sykdom eller andre utfordringer. Testene underveis er derfor ikke inkludert i analysene for at det skulle være et større utvalg i analysene. Derfor er resultatene før og etter treningsintervensjonen det som er tatt med i analysene.

Kroppssammensetning for bestemmelse av mager muskelmasse er bestemt ved dual-energy X-ray absorptiometry (DXA) før og etter intervensjonen (Lunar Prodigy, GE Healthcare, Oslo, Norway). Før DXA-målinger fikk deltakerne beskjed om å faste 2 timer og avstå fra krevende fysisk aktivitet i 48 timer. Det var også minimum 48 timer fra siste styrkeøkt til DXA-måling.

## 5.2.4 Dataanalyser og statistikk

Statistiske analyser er gjort i R studio<sup>21</sup>. Det er gjort enkle lineære regresjonsmodeller på differansen mellom gruppene (ett sett & flere sett) i endring av styrke og muskelmasse i løpet av intervensjonen. For maksimal styrke er det gjort analyser på øvelsene beinpress og kneekstensjon. Muskelmasse er målt som endringen i mager muskelmasse i beinet som har trent ett mot beinet som har trent tre sett.

## 5.3 Resultater

DXA-resultatene viste at gjennomsnittlig differanse mellom ett og tre sett var 122.79 (95 % KI: [8.59-237],  $p = 0.04$ ). Også for styrkeøvelsene var forbedringen i 1RM i gjennomsnitt større for det beinet som hadde trent flere sett. I beinpress var forskjellen 7.22 (95 % KI: [0.9-13.5],  $p = 0.026$ ), mens for kneekstensjon var det 3.6 (95 % KI: [1.4-5.8],  $p = 0.002$ ) differanse.

Tabellen nedenfor viser nivået ved baseline for styrkeøvelsene og mager muskelmasse.

Table 5.1: Resultater fra pre-test

Volum	Magermasse (g)	Beinpress (kg)	Kneekstensjon (kg)
multiple	8,603.5 $\pm$ 2,032.9	208.1 $\pm$ 76.4	69.2 $\pm$ 23.3
single	8,589.0 $\pm$ 2,021.0	217.9 $\pm$ 76.1	74.3 $\pm$ 25.5

*Data er presentert som gjennomsnitt  $\pm$  standardavvik.*

Figuren viser om det er en sammenheng mellom prosentvis endring i muskelmasse og 1RM beinpress. 0.5 % av endringen i 1RM beinpress kan forklares med endringen i muskelmasse ( $R = 0.005$  &  $p = 0.59$ ).

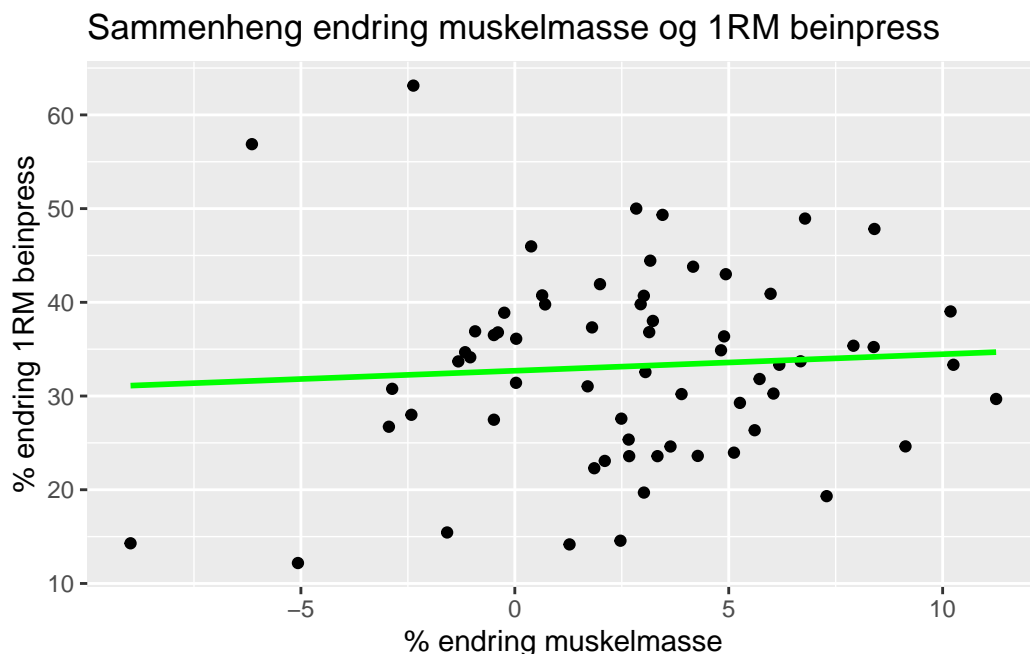


Figure 5.1: Figur som viser en lineær regresjonsmodell med endring i muskelmasse som prediktor variabel for endring i 1RM beinpress.

## 5.4 Diskusjon

I det beinet som trente tre sett, så man at det var mer hypertrofi og større økning i styrke, målt i de to øvelsene, sammenlignet med de som trente ett sett. Prosentvis endring i magermasse var høyere ved 3 sett, henholdsvis  $3.1 \pm 4.4$  % mot ett sett, som økte  $1.9 \pm 3.5$  %. Man har sett det samme i en metaanalyse, som så at to til tre sett og fire til seks sett ga bedre resultat på effektstørrelse enn ett sett (ingen forskjell mellom to til tre og fire til seks sett)<sup>22</sup>. Vi skal være litt forsiktige med å trekke konklusjoner basert på metaanalyser, men mange andre studier sammenligner grupper med forskjellige deltakere og ikke samme deltaker med ulikt treningsvolum på to forskjellige bein<sup>23</sup>. Dette gjør at man ikke får tatt høyde for biologiske variasjoner hos de ulike individene, dermed blir det vanskelig å sammenligne ulikt volum. Økningen i muskelmasse kan i analysene fra denne rapporten ikke forklare økningen i 1 RM beinpress når begge bein ble inkludert i korrelasjonsanalysen.

Til tross for at det ikke var noen sammenheng, så man også en større økning i styrke, i beinpress ved tre sett kontra ett sett (den prosentvise endringen på henholdsvis  $34.4 \pm 10.5$  &  $31.9 \pm 10$ ). Man så det samme på kneekstensjon, tre sett økte her  $32.5 \pm 9.1$  %, ett sett derimot økte med  $30.4 \pm 8.8$  %.

Som konklusjon ut i fra de gjennomførte analysene fra rapporten, kan vi si at responsen på styrke og hypertrofi følger et dose-volum forhold. Dette vil si at tre sett er mer gunstig enn ett sett for ikke-røykende, utrente kvinner og menn mellom 18 og 40 år.

## 6 Vitenskapsteori

Induksjon er en prosess som blir brukt for å trekke konklusjoner basert på observasjoner vi har gjort<sup>24</sup>. Metoden, eller prosessen blir gjerne brukt for å trekke generelle konklusjoner angående hvordan verden er satt sammen og hvordan den fungerer. Konklusjonene baserer seg på spesifikke observasjoner som har blitt gjort tidligere, og observasjonene har gjerne blitt testet om igjen, og om igjen for å sørge for at ting ikke har skjedd ved én tilfeldighet.

Det finnes både fullstendig og ufullstendig induksjon. Et eksempel på fullstendig induksjon er ved valg, valgresultatet baserer seg på induksjon. Alle stemmesedlene blir telt opp og vi kan konkludere med at kandidaten som har fått flest stemmer, vinner valget. Ved fullstendig induksjon kan vi være helt sikre på at konklusjonen vi har kommet frem til er rett, siden vi har sett på alle stemmesedlene. Ved ufullstendig induksjon baserer man seg på et begrenset antall observasjoner og trekker en generell konklusjon deretter. Ved ufullstendig induksjon vil det alltid være en fare for at konklusjonen er usann, siden man ikke undersøker alle relevante tilfeller.

Hume mener på sin side at induksjon rett og slett ikke er så mye verdt. I Humes' øyne er det ingen logisk forklaring til at fremtiden vil fortsette å være lik fortiden i all fremtid. Erfaringer vi har gjort oss tidligere sier heller ikke noe om fremtiden, selv om det gir en god pekepinn. Selv om to biler krasjer i et bestemt kryss i dag, betyr ikke det at to biler vil krasje i det samme krysset i morgen. Hume konkluderer med at induksjon kun er basert på vaner og/eller forventninger og at det ikke er basert på rasjonell grunn.

En innvendig mot Humes premiss er at vi mennesker klarer å gjøre gode resonnementer, gjenkjenne mønstre og tenke logisk. Med bakgrunn i dette kan vi si at våre forventninger om fremtiden ikke kun er basert på forventninger og vaner, men fra en dypere rasjonell forståelse av verden. At ting har blitt som de har, er et resultat av den stadige evolusjonen og utviklingen vi har hatt som menneskehet. Vi og verden har utviklet oss for å stadig bedre vår overlevelsessevne. Og dermed vil vi også fortsette å utvikle oss videre.

Hume ville i dette tilfellet antakelig svart med noe som at; selv om vi i all tid så langt har utviklet oss som menneskehet, og verden stadig har tatt fremskritt, betyr det ikke at vi vil fortsette med det i morgen. Selv om all fornuft tyder på at morgendagen vil ligne på dagen i går, siden dagen før der lignet på dagen i går, betyr det ikke at det vil være sann. Selv om vi mennesker har blitt gode på å kjenne igjen mønstre, kan man fortsatt ikke forutse hvordan morgendagen blir. Det vet vi kun i morgen.

Hovedessensen av prinsippet om falsifikasjon, eller falsifikasjonisme handler om at vi i vitenskapen ikke bør prøve å finne argumenter for at noe stemmer, men at vi heller bør prøve å falsifisere, altså bevise at noe er usant. Karl Popper var opptatt av at teorier skulle kunne motbevises ved hjelp av eksperimenter, eller observasjoner. Et eksempel vi kan bruke fra idretten er teorien om at bruken av høydetrening kan forbedre utholdenhetsprestasjonen. Hvis man har to forskjellige grupper, en på høydetrening og en uten høydetrening. Og det ikke ses noen forskjell i utholdenhetsprestasjon, kan hypotesen om høydetrening falsifiseres. Hvis teorier kommer seg gjennom mange forsøk på falsifikasjon, kan vi se på den som at den foreløpig er korrekt, men aldri som endelig bekreftet. Popper

gikk på sin side så langt, som å si at søken etter bekreftelser på hypoteser er selve kjennetegnet på pseudovitenskap.

Et problem med falsifikasjonisme, hentet fra idretten er effekten av varmetrening på utholdenhetsprestasjon. Det har i dag blitt vanligere og vanligere å drive med varmetrening, enten i oppvarmet rom eller ved bruk av varmedrakt (for eksempel ullundertøy under regntøy). Det er mye diskutert hva slags effekt dette kan ha på utholdenhetsprestasjonen, også i konkurranser eller tester hvor det er normale forhold (hverken oppvarmet rom eller varmedrakt i bruk). Det er en del andre variabler som spiller inn her, spesielt individuelle forskjeller som gjøres blant utøvere både underveis og i ettertid av en varmetreningsøkt. Eksempler på dette er hydrering underveis og i etterkant, hvile og jern som kosttilskudd. Dette gjør det vanskelig å teste denne hypotesen på en falsifiserbar måte. En mulig løsning på dette problemet er å gjøre lange kontrollerte studier der utøvere gjennomgår trening i varme forhold (varmekammer/varmedrakt), med en stor nok kontrollgruppe som ikke gjør noen form for varmetrening. I tillegg bør man i dette tilfellet sørge for kontrollert oppfølging av både hydrering (sørge for at deltakerne drikker lik mengde), hvile og andre faktorer som kan påvirke effekten av varmetreningen. Dette er ulike tiltak som gjør at hypotesen blir mer falsifiserbar.

Et annet eksempelproblem innenfor idrettsforskning og falsifikasjonisme er som følger. Det settes opp et prosjekt som skal se på utvikling av utholdenhet og styrke hos en rekke deltakere. Det gjøres baselinetester før en treningsintervensjon på 12 uker. Etter de 12 ukene med trening er gjennomført gjennomføres det nye tester for å se hvordan deltakerne nå presterer. Halve gruppen har trent harde intervaller, og den andre halvdelens har trent rolig trening. Gruppen som har gjennomført intervalltrening har oppnådd mest fremgang, men problemet er at vi ikke har tatt høyde for individuelle forskjeller blant deltakerne som er med. En mulig løsning på dette problemet er at vi lar deltakerne få en 12 ukers washout-periode hvor det ikke blir gjennomført noe trening. Man gjør nye baselinetester for å se om deltakerne nå har mistet treningsadaptasjonene de opparbeidet seg i forkant. Vi gjennomfører nå 12 nye uker med trening, og deltakerne har nå byttet om, de som trente intervalltrening under forrige intervensjon får nå rolig trening denne gangen. På denne måten tar vi høyde for individuelle forskjeller, ettersom vi nå har byttet om på deltakerne. Dette er en måte å gjøre hypotesen mer falsifiserbar på.

Et tredje forsøk fra idrettsverden er placeboeffekt, det finnes mye forskning gjort på kosttilskudd innenfor idretten. Hvis en utøver tror man får et kosttilskudd som skal gi en forventet bedring i prestasjon kan dette være nok til at man presterer bedre. Dette gjør det vanskelig å falsifisere hypoteser. En mulig løsning på dette problemet er å gjøre double blinded eller dobbeltblinda eksperimenter, så hverken forsker eller forsøksperson vet om man får placebo eller om man får det faktiske kosttilskuddet. Dette gjør hypotesen mer falsifiserbar, for hvis man i dette tilfellet også finner effekt, kan man med ganske god sannsynlighet si at det er på grunn av kosttilskuddets virkemiddel og ikke at utøveren tror man har fått tilskuddet.

# 7 Lab-rapport

## 7.1 Introduksjon

Genekspresjonsanalyse med kvantitativ fluorescens-basert sanntids polymerasekjedereaksjon (qPCR) er en avansert forskningsmetode som ofte anvendes innen treningsfysiologi. Målet med qPCR er å studere genuttrykk for et spesifikt målgen ved bruk av biologisk materiale<sup>25</sup>. qPCR er særlig nyttig for å måle treningsinduserte endringer i genuttrykk i muskelfibertyper. Den største forskjellen mellom qPCR og tradisjonell PCR er at qPCR gjør det mulig å kvantifisere og måle mengden av DNA-sekvenser i sanntid<sup>25</sup>.

For å utføre qPCR kreves RNA, som ekstraheres fra biologisk materiale, for eksempel muskelvev. RNA-et må gjennom flere trinn før det kan brukes i analysen. Det første trinnet er å omdanne RNA til komplementært DNA (cDNA) ved hjelp av en prosess som kalles revers transkripsjon<sup>25</sup>. Deretter kopieres cDNA eksponentielt, slik at milliarder av kopier kan fremstilles<sup>25</sup>.

## 7.2 Metode

Ved oppstart av forsøket fikk vi utdelt ferdigprodusert cDNA fra et tidligere gjennomført styrketreningsprosjekt, levert av labansvarlig. For å gjennomføre qPCR ble det brukt cDNA kombinert med en Master Mix, som inneholdt følgende komponenter:

SYBR Green Mix: 50 µl Vann (H<sub>2</sub>O): 20 µl Primer-mix (enten MHC1, MHC2x eller MHC2a): 10 µl I tillegg ble det laget en kontroll-Master Mix som besto av følgende:

b2m Primer-mix: 50 µl Vann (H<sub>2</sub>O): 100 µl SYBR Green Mix: 250 µl Fremstilling av fortynningsrekker

Under forsøket ble det også utarbeidet to ulike fortynningsrekker for å standardisere og validere resultatene. Fortynningsrekkene var som følger:

Fortynningsrekke 1: 1 (ufortynnet) 1/10 1/100 1/1000 Fortynningsrekke 2: 1/2 1/20 1/200 For å lage fortynningsrekkene ble c-myc-primer fortynnet med H<sub>2</sub>O. I den ufortynnede prøven (1/1) ble det tilsatt 30 µl av prøven og 0 µl H<sub>2</sub>O, mens de resterende fortynningene ble fremstilt ved proporsjonal tilsetning av vann og prøve.

Dette oppsettet sikret nøyaktighet og reproducerbarhet i analysen, og tillot oss å undersøke genuttrykket under forskjellige fortynningsforhold.

Table 7.1: Fortynning Tabell

Fortynning	Prøve	H2O
1	30	0

Fortynning	Prøve	H2O
1/10	2	18
1/100	2	18
1/1000	2	18
½	10	10
1/20	2	18
1/200	2	18

Flytter 2 µl fra rør 1 til 2a, og 10µl fra 1 til 2b, vortex rør 2a+2b.

Flytter 2µl fra 2a til 3a og 2µl fra 2b til 3b, vortex rør 3a+3b.

Flytter 2µl fra 3a til 4a og 2µ fra 3b til 4b, vortex rør 4a + 4b.

Table 7.2: Tabell: Oversikt over prøver

	Pre_week_0	Post_week_12	cmymc_1	cmymc_2a	cmymc_3a	cmymc_4a	cmymc_2b	cmymc_3b	cmymc_4b
A	myhe 1	myhe 1	NA	NA	NA	NA	NA	NA	NA
B	myhe 1	myhe 1	NA	NA	NA	NA	NA	NA	NA
C	myhe 1	myhe 1	NA	NA	NA	NA	NA	NA	NA
D	myhe 2a	myhe 2a	NA	NA	NA	NA	NA	NA	NA
E	myhe 2a	myhe 2a	NA	NA	NA	NA	NA	NA	NA
F	myhe 2a	myhe 2a	NA	NA	NA	NA	NA	NA	NA
G	myhe 2x	myhe 2x	NA	NA	NA	NA	NA	NA	NA
H	myhe 2x	myhe 2x	NA	NA	NA	NA	NA	NA	NA
I	myhe 2x	myhe 2x	NA	NA	NA	NA	NA	NA	NA
J	b2m	b2m	NA	NA	NA	NA	NA	NA	NA
K	b2m	b2m	NA	NA	NA	NA	NA	NA	NA
L	b2m	b2m	NA	NA	NA	NA	NA	NA	NA

Prøvene ble pipettert over i brønner i henhold til et forhåndsdefinert pipetteringskart. Hver brønn ble fylt med 8 µl primer-spesifikk prøve og 2 µl cDNA-løsning eller kontrolløsning (A-L, 1 og 2). Fortynningsrekkene ble pipettert inn i brønnene 5-11 (A-C). Alle prøvene ble satt opp i et treplatt-format, noe som sikret nøyaktighet og reproducerbarhet i analysen.

qPCR-analysen ble gjennomført ved hjelp av Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies AS) og programvaren Quant Studio 5. Prosessen besto av tre hovedtrinn. Først ble prøvene kjørt i en «Hold stage» der temperaturen økte med 1,99 °C/s opp til 50 °C, hvor den ble holdt konstant i 2 minutter. Deretter økte temperaturen videre med 1,99 °C/s opp til 95 °C, hvor den ble holdt i ytterligere 2 minutter.

Selve PCR-prosessen, kalt «PCR stage», besto av 40 sykluser. Hver syklus innebar at temperaturen først ble holdt på 95 °C i 1 sekund før den senket seg med 1,77 °C/s til 60 °C. Denne temperaturen ble deretter holdt konstant i 30 sekunder. Etter hver syklus ble det tatt et bilde av fluorescensen for å overvåke reaksjonen i sanntid.

Til slutt ble prøvene kjørt gjennom en «Melt stage». Her økte temperaturen med 1,99 °C/s opp til 95 °C, hvor den ble holdt i 15 sekunder. Deretter ble temperaturen gradvis senket med 1,77 °C/s til 60 °C, hvor den ble holdt konstant i 1 minutt. Til slutt økte temperaturen med 0,15 °C/s opp til 95

°C igjen, hvor den ble holdt i 15 sekunder. Etter dette var PCR-prosessen ferdig, og CT-verdiene ble hentet ut for videre analyse.

### 7.2.1 Databehandling

De innsamlede CT-verdiene ble analysert og behandlet ved hjelp av Microsoft Excel. Dette inkluderte utregninger og datavisualisering for å tolke genuttrykk og endringer i fluorescens. Hele prosessen sikret presise resultater og pålitelige konklusjoner.

Table 7.3: Ct-verdier for de ulike prøvene

Sample Name	Target Name	Ct Data			
		Ct1	Ct2	Ct3	Average Ct
Control sample W0	b2m	23.427	24.072	23.318	23.60567
W0	MHC I	18.299	19.223	19.764	19.09533
W12	MHC I	18.431	19.080	18.348	18.61967
W0	MHC 2a	22.419	17.707	18.314	19.48000
W12	MHC 2a	18.393	18.731	35.236	24.12000
W0	MHC 2x	25.708	25.036	23.771	24.83833
W12	MHC 2x	25.120	23.575	23.052	23.91567

Tallene viser til antall sykluser før syklisk terskel (CT) er nådd. CT verdiene verdier for å nå syklisk terskel har endret seg fra uke 0 til uke 12. For MHC1 har det endret seg fra et gjennomsnitt på 19,09 -18,62 sykluser. For MHC2a har det endret seg fra 19,48 -24,12 og for MHC 2x fra 24,83-23,91. Et lavere antall sykluser et større genuttrykk.

Uke	Prosentvis fordeling av genuttrykk		
	MHC1	MHC2a	MHC2x
<b>Uke 0</b>	56	43	1
<b>Uke 12</b>	95	2	2

Vi ser at mengden genuttrykk for muskelfibertype 2a (MHC-2a) har sunket fra 43% til 2% etter 12 uker, det har også vært en reduksjon for muskelfibertype 2x (MHC-2x) fra 2% til 1%. Muskelfibertype 1 (MHC1) har derimot økt fra 56% til 95%.

Table 7.5: Primer Efficiency Results

Slope	Primer Efficiency
<b>-2.477405</b>	153.31



## 7.3 Diskusjon

Målet med dette forsøket var å undersøke endringer i myoisintungkjedene. Etter en styrketreningsintervensjon på utrent forsøksperson på 12 uker, ved hjelp av fluoricens- basert sanntids kvantitativ polymerase kjedereaksjon (qPCR).

Det har i dette forsøket blitt undersøkt hvor mange sykluser CT-verdiene til den ulike myoisintungkjedene trenger for å nå sin terskelverdi. Der færre sykluser og lavere CT verdier indikerer et større genuttrykk. I våre resulatater har vi sett en endring for alle myosintungkjedene når det gjelder antall sykluser myoisnkjedene har trengt for å nå sin CT-verdi og den prosentvise endringen for tungkjedene indikerer at vi har fått en stor endring fra 56%- 95% for MHC1, for MHC2a har det blitt redusert fra 43% til 2%, det samme gjelder for MHC2x som har fått en reduksjon fra 2%-1%.

I en tidligere studie av Ellefsen et al. (2014), hvor en styrketreningsintervensjon ble gjennomført på utrente individer over 12 uker, ble det observert en økning i MHC2A, en reduksjon i MHC2X, samt stabilitet i MHC1<sup>26</sup>. I kontrast til dette viser våre resultater motstridende funn, med både reduksjon i MHC2A og MHC2X, samt en betydelig økning i MHC1. Andre studier, som<sup>27</sup> Terzis et al. (2008) og<sup>28</sup> også vist at utrente personer med overvekt av MHC2X opplever en reduksjon i MHC2X og en økning i MHC2A ved trening, med minimal endring i MHC1. Det er kjent at genuttrykk ikke kan endres fra MHC1 til MHC2A eller MHC2X, noe som gjør det vanskelig å forklare de resultatene vi har fått fra analysen av myosintungkjeder. Dette reiser spørsmål om hva som kan ha skjedd under vår analyse og om det er spesifikke faktorer ved vårt eksperiment som kan ha bidratt til disse avvikene fra tidligere forskning.

Det som en kan reise en usikkerhet om er om både pippetringferdigheter og primerkvaliteten har vært god nok enten at primer har gått ut på dato eller bruk av feil primer. Primer efficiency skulle ligge mellom 90-110%. I vårt tilfelle fikk vi en efficiency på 153% noe som kan tyde på at det kan være menneskelige feil som har påvirket resultatet. Det er også vanskelig å skulle trekke slutninger basert på en prøve. Samt at vi har ingen forkunnskap om hvilket treningsstimuli forsøkspersonen vi har fått har vært utsatt for, annet enn den informasjonen vi har fått fra labansvarlig.

### 7.3.1 Konklusjon

Basert på funn vi har fått i forsøket kan vi ikke si noe om endringene i myosintungkjedene for denne forsøkspersonen. Da resultatene vi har fått ikke er mulig basert på det vi kjenner til av tidligere forskning.

# Referanser

1. Hopkins, W. G. (2000). Measures of reliability in sports medicine and science [Journal Article]. *Sports Med*, 30(1), 1–15. <http://www.ncbi.nlm.nih.gov/pubmed/10907753>
2. Spiegelhalter, D. (2020). Introducing the art of statistics: How to learn from data. *Numeracy*, 13(1).
3. Bassett, D. R., Jr, & Howley, E. T. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med. Sci. Sports Exerc.*, 32(1), 70–84.
4. Joyner, M. J., & Coyle, E. F. (2008). Endurance exercise performance: The physiology of champions. *J. Physiol.*, 586(1), 35–44.
5. Tanner, R. K., & Gore, C. J. (2012). *Physiological tests for elite athletes 2nd edition* [Book]. Human Kinetics. <https://books.google.no/books?id=0OPIiMks58MC>
6. Halperin, I., Pyne, D. B., & Martin, D. T. (2015). Threats to internal validity in exercise science: A review of overlooked confounding variables. *Int. J. Sports Physiol. Perform.*, 10(7), 823–829.
7. Pisciă, D., Dammers, R., Boersma, E., & Volovici, V. (2022). Tenets of good practice in regression analysis. A brief tutorial. *World Neurosurg.*, 161, 230–239.e6.
8. Machado, F. A., Nakamura, F. Y., & Moraes, S. M. F. D. (2012). Influence of regression model and incremental test protocol on the relationship between lactate threshold using the maximal-deviation method and performance in female runners. *Journal of Sports Sciences*, 30(12), 1267–1274. <https://doi.org/10.1080/02640414.2012.702424>
9. Rønnestad, B. R., Hansen, J., & Ellefsen, S. (2014). Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists. *Scand. J. Med. Sci. Sports*, 24(1), 34–42.
10. Rønnestad, B. R., Ellefsen, S., Nygaard, H., Zacharoff, E. E., Vikmoen, O., Hansen, J., & Hallén, J. (2014). Effects of 12 weeks of block periodization on performance and performance indices in well-trained cyclists. *Scand. J. Med. Sci. Sports*, 24(2), 327–335.
11. Breil, F. A., Weber, S. N., Koller, S., Hoppeler, H., & Vogt, M. (2010). Block training periodization in alpine skiing: Effects of 11-day HIT on VO<sub>2</sub>max and performance. *Eur. J. Appl. Physiol.*, 109(6), 1077–1086.
12. Rønnestad, B. R., Hansen, J., Thyli, V., Bakken, T. A., & Sandbakk, Ø. (2016). 5-week block periodization increases aerobic power in elite cross-country skiers. *Scand. J. Med. Sci. Sports*, 26(2), 140–146.
13. Solli, G. S., Tønnessen, E., & Sandbakk, Ø. (2019). Block vs. Traditional periodization of HIT: Two different paths to success for the world’s best cross-country skier. *Front. Physiol.*, 10, 375.
14. Carpinelli, R. N., & Otto, R. M. (1998). Strength Training: Single Versus Multiple Sets. *Sports Medicine*, 26(2), 73–84. <https://doi.org/10.2165/00007256-199826020-00002>
15. Sooneste, H., Tanimoto, M., Kakigi, R., Saga, N., & Katamoto, S. (2013). Effects of Training Volume on Strength and Hypertrophy in Young Men. *Journal of Strength and Conditioning Research*, 27(1), 8–13. <https://doi.org/10.1519/JSC.0b013e3182679215>
16. Radaelli, R., Fleck, S. J., Leite, T., Leite, R. D., Pinto, R. S., Fernandes, L., & Simão, R. (2015). Dose-Response of 1, 3, and 5 Sets of Resistance Exercise on Strength, Local Muscular

- Endurance, and Hypertrophy. *Journal of Strength and Conditioning Research*, 29(5), 1349–1358. <https://doi.org/10.1519/JSC.0000000000000758>
17. Cannon, J., & Marino, F. E. (2010). Early-phase neuromuscular adaptations to high- and low-volume resistance training in untrained young and older women. *Journal of Sports Sciences*, 28(14), 1505–1514. <https://doi.org/10.1080/02640414.2010.517544>
  18. Mitchell, C. J., Churchward-Venne, T. A., West, D. W. D., Burd, N. A., Breen, L., Baker, S. K., & Phillips, S. M. (2012). Resistance exercise load does not determine training-mediated hypertrophic gains in young men. *Journal of Applied Physiology*, 113(1), 71–77. <https://doi.org/10.1152/jappphysiol.00307.2012>
  19. Marx, J. O., Ratamess, N. A., Nindl, B. C., Gotshalk, L. A., Volek, J. S., Dohi, K., Bush, J. A., Gomez, A. L., Mazzetti, S. A., Fleck, S. J., H??Kkinen, K., Newton, R. U., & Kraemer, W. J. (2001). Low-volume circuit versus high-volume periodized resistance training in women: *Medicine and Science in Sports and Exercise*, 635–643. <https://doi.org/10.1097/00005768-200104000-00019>
  20. Messier, S. P., & Dill, M. E. (1985). Alterations in Strength and Maximal Oxygen Uptake Consequent to Nautilus Circuit Weight Training. *Research Quarterly for Exercise and Sport*, 56(4), 345–351. <https://doi.org/10.1080/02701367.1985.10605339>
  21. Posit team. (2023). *RStudio: Integrated development environment for r*. Posit Software, PBC. <http://www.posit.co/>
  22. Krieger, J. W. (2009). Single Versus Multiple Sets of Resistance Exercise: A Meta-Regression. *Journal of Strength and Conditioning Research*, 23(6), 1890–1901. <https://doi.org/10.1519/JSC.0b013e3181b370be>
  23. Starkey, D. B., Pollock, M. L., Ishida, Y., Welsch, M. A., Brechue, W. F., Graves, J. E., & Feigenbaum, M. S. (1996). Effect of resistance training volume on strength and muscle thickness: *Medicine & Science in Sports & Exercise*, 28(10), 1311–1320. <https://doi.org/10.1097/00005768-199610000-00016>
  24. Thurén, T. (2022). *Vitenskapsteori for nybegynnere*.
  25. Kuang, J., Yan, X., Genders, A. J., Granata, C., & Bishop, D. J. (2018). An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PLOS ONE*, 13(5), e0196438. <https://doi.org/10.1371/journal.pone.0196438>
  26. Ellefsen, S., Vikmoen, O., Zacharoff, E., Rauk, I., Slettaløkken, G., Hammarström, D., Strand, T. A., Whist, J. E., Hanestadhaugen, M., Vegge, G., Fagernes, C. E., Nygaard, H., Hollan, I., & Rønnestad, B. R. (2014). Reliable determination of training-induced alterations in muscle fiber composition in human skeletal muscle using quantitative polymerase chain reaction. *Scandinavian Journal of Medicine & Science in Sports*, 24(5). <https://doi.org/10.1111/sms.12185>
  27. Terzis, G., Georgiadis, G., Stratakis, G., Vogiatzis, I., Kavouras, S., Manta, P., Mascher, H., & Blomstrand, E. (2008). Resistance exercise-induced increase in muscle mass correlates with p70S6 kinase phosphorylation in human subjects. *European Journal of Applied Physiology*, 102(2), 145–152. <https://doi.org/10.1007/s00421-007-0564-y>
  28. Andersen, J. L., & Aagaard, P. (2000). Myosin heavy chain IIX overshoot in human skeletal muscle. *Muscle & Nerve*, 23(7), 1095–1104. [https://doi.org/10.1002/1097-4598\(200007\)23:7%3C1095::AID-MUS13%3E3.0.CO;2-O](https://doi.org/10.1002/1097-4598(200007)23:7%3C1095::AID-MUS13%3E3.0.CO;2-O)