

Study design

Kandidatnummer

2022-12-02

Table of contents

Preface	4
Deloppgave 1: Beskrivende statistikk, reliabilitet og validitet, verktøy for reproduserbar dataanalyse	5
Introduksjon	5
Metode	6
Praktisk gjennomføring/tiltak for å sikre reliabilitet	7
Resultater	7
Deloppgave 2: Vitenskapsfilosofi	10
Introduksjon	10
Del 1	10
Beskrivelse og forklaring av falsifikasjonisme	10
Problemet med falsifikasjonisme, Poppers og Duhems læsning på problemet . .	11
Del 2	12
Bayesianisme og Bayes Theorem	12
HD metoden	13
Adbuktivisme	14
HD-teori og Abduktivisme vs Bayesianisme	14
Utfordringen med Bayesianisme og evaluering av hvor alvorlig problemet er . .	14
Deloppgave 3: Regresjonsmodeller	16
Oppgave 1	16
Undersøke treningsintensiteten på 2mmol L-1 og 4mmol L-1 hos forsøksperson 10 og 3	16
Laktatterskel på forsøksperson 10 etter Machodo “lactate threshold estimated by the maximal deviation method (LTDmax)”.	19
Laktatterskel på forsøksperson 3 etter Machado “lactate threshold estimated by the maximal deviation method (LTDmax)”	20
“Predicted - fixed lactate value”	20
“Predicted - fixed lactate value”	20
Endring i laktat 125 watt og 250 watt	21
Oppgave 2	22
Oppgave 3	24
Deloppgave 4: Statistisk interferens	27

Deloppgave 5: Studiedesign	35
Styrketrening for eldre	35
Introduksjon	35
Spørsmål:	37
Sekundære spørsmål	37
Design	37
Statistiske analyser	40
Resultat	41
Konklusjon	42
References	43

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

```
1 + 1
```

```
[1] 2
```

Deloppgave 1: Beskrivende statistikk, reliabilitet og validitet, verktøy for reproduserbar dataanalyse

Introduksjon

Vi gjennomførte to testdager 26.09.2023 og 28.09.2023 i tiden 08:00-16:00. Hensikten med disse to dagene var å gjennomføre fysiologiske tester med høy grad av reliabilitet. Det er flere faktorer som påvirker både validitet og reliabilitet, og det er viktig å ta høyde for dette under fysiologisk testing. Vi gjennomførte testdag 1 og testdag 2 med kun én dag mellom for å sikre at deltakerne var på tilsvarende fysiologisk nivå ved begge testene. Vi tok derfor en rekke forhåndsregler for å sikre så like testforhold som mulig.

Reliabilitet refererer til reproduserbarheten til en f.eks. en fysiologisk test som gjennomføres flere ganger i en repetert studie, der bedre reliabilitet indikerer bedre presisjon og måling av endring over tid (**hopkinsMeasuresReliabilitySports2000?**). Innenfor reliabilitet er det en rekke relevante begreper. Standardavvik (SD) forteller hvor langt unna dataene er fra gjennomsnittet (**spiegelhalter2020?**), typical error (TE) beskrives av (**hopkinsMeasuresReliabilitySports2000?**) som variabiliteten hos hver enkelt verdi og tenkes å kunne visualisere feilmarginen av et estimat. For å få nøyaktige målinger som kna sammenliknes er det viktig med presise måleinstrumenter som kalibreres nøye. Det er også faktorer som læringseffekt, motivasjon, restitusjon og ernæringstilstand som

kan påvirke resultatene, og det er viktig å ta høyde for dette ved fysiologisk testing. (hopkinsMeasuresReliabilitySports2000?) hevder at det kreves om lag 50 deltakere og 3 repeterte målinger for å kunne estimere reliabiliteten. Dette er for å utelukke de overnevnte faktorene.

Kroppens maksimale oksygenopptak (VO_{2maks}) gir informasjon om en persons maksimale aerobe kapasitet. Oksygenopptaket bestemmes av både sentrale- og perifere faktorer og kan illustreres ved Flick's likning:

$$VO_2 = (HR \times SV) \times (aO_2 - vO_2)$$

En VO_{2maks} -test går ut på at man måler hvor mange ml oksygen en person evner å ta opp og omsette per minutt. Oksygenkravet øker lineært med belastningen helt til personen når sin maksimale aerobe kapasitet, da vil kurven flate ut eller eventuelt synke. En persons maksimale oksygenopptak kan beskrives både i form av absolutte tall (ml/min) eller som relative tall i forhold til kroppsvekt (ml/kg/min).

Metode

VO_{2maks} testen gjennomføres som en trappetrinnstest der motstanden øker med 25W hvert minutt til utmattelse/når $RPM < 60$. VO_2 målinger registreres hvert 30 sek. Deltakerne startet testen på enten 150W, 200W eller 250W avhengig av fysisk form og erfaring med sykkel. Hvert minutt øker watten med 25 helt til utmattelse. Etter endt test ble informasjon innhentet og plottet i ferdigstilt Excel-dokument.

Praktisk gjennomføring/tiltak for å sikre reliabilitet

Selv om en tydelig protokoll er essensielt for å sikre reliable tester på en fysiologilab, er det flere hensyn som må tas underveis. Vi begynte hver test med å kalibrere utstyret slik at det var oppdatert etter forholdene til hver deltaker hver klokke time, for å minimere risikoen for at utstyret skal måle feil (**tannerPhysiologicalTestsElite2013?**). For å sikre lik grad av verbal motivasjon og formulering av instruks, valgte vi å bruke samme testleder på hver person (**halperinThreatsInternalValidity2015a?**). Vi ha også instruks om at kosthold og søvn skulle være likt før og på testdagen, samt ingen trening mellom testene da dette er faktorer som kan påvirke metabolisme og prestasjon, dette ble dog ikke kontrollert (**tannerPhysiologicalTestsElite2013?**).

Resultater

Etter testdag 1 (T1) fikk vi en oversikt hvilket fysiologisk nivå deltakerne i prosjektet vår på. Vi valgte å undersøke variablene VO_{2max} (ml/min), VO_{2max} (ml/min/kg) og kroppsvekt. Se Table 1 for oversikt over T1.

Tabell 1: Oversikt over testresultater etter T1. Gjennomsnitt (Mean), laveste observasjon (Min), høyeste observasjon (Max), og standardavvik (SD)

Variable	Mean	Min	Max	SD
VO2max (ml/min)	3,990.12	2,504.00	5,893.00	999.95
VO2max (ml/min/kg)	49.83	31.74	77.54	15.28
Weight (kg)	81.75	66.40	105.90	12.32

Formålet med dette prosjektet var å teste reliabiliteten på utvalgte fysiologiske mål mellom T1 og T2. Figure 1 gir en oversikt over resultatet på T1 og T2 for hver ID. Oversikt over aktuelle reliabilitetstall finner man i Table 2.

Standardavvik (SD) kan forklares ved et mål på hvor stor spredningen er i forhold til datapunktene's middelvei/gjennomsnitt, og definerer hvert enkelt datapunkts avvik fra gjennomsnittet (spiegelhalterArtStatisticsLearning2020?).

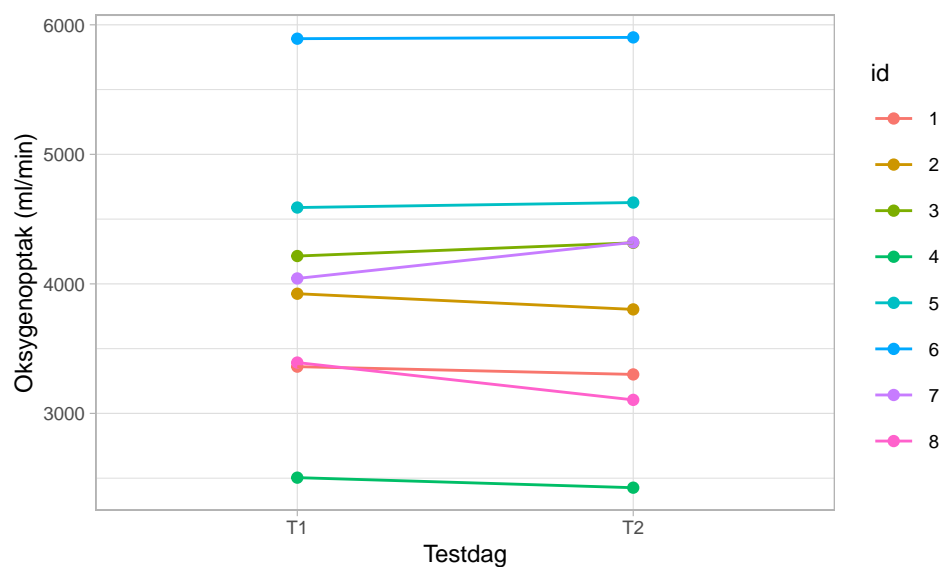
Variasjonskoeffisienten (CV) angir et spredningsmål for verdiene i et datasett. CV uttrykker ofte variasjon i forhold til gjennomsnittsverdien og angis i prosent (spiegelhalterArtStatisticsLearning2020?).

Typical Error (TE) eller standardfeil er variabiliteten hos hver enkelt verdi og tenkes å kunne visualisere feilmarginen av et estimat. Eksempelvis vil en standardfeil kunne forklares gjennom biologiske prosesser som påvirker for eksempel kraftutvikling, som følge av mentale eller fysiske faktorer (hopkinsMeasuresReliabilitySports2000?).

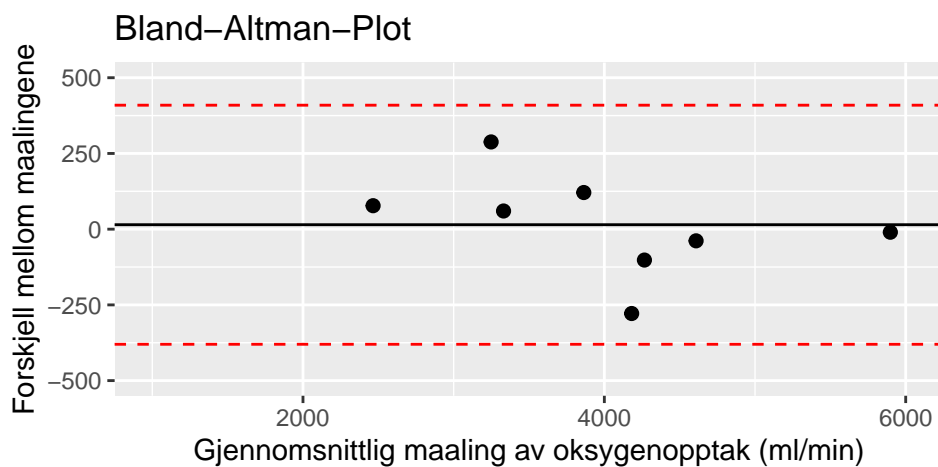
Limits of agreement (LoA) viser til 1,96 standardavvik fra den gjennomsnittlige differansen mellom T1 og T2. Dette er illustrert i Figure 2 med et Bland-Altman-Plot.

Tabell 2: Oversikt over reliabilitetstall for utvalgte tester. Tabellen viser gjennomsnitt av T1 og T2 (Mean), standardavvik (SD), typical error (TE), variasjonskoeffisient (CV), og limits of agreement (LoA)

Variable	Mean	SD	TE	CV (%)	LoA
VO2max (ml/min)	3,982.78	166.91	118.02	2.96	394.68
VO2max (ml/min/kg)	49.50	2.48	1.75	3.54	5.87
Weight (kg)	82.07	1.73	1.22	1.49	4.09



Figur 1: Sammenligning av oksygenopptak målt i ml/min mellom testdag 1 og testdag 2 for hver forsøksperson



Figur 2: Bland-Altman-Plot for maksimalt oksygenopptak (ml/min). Stiplede linjer tilsvare øvre og nedre limits of agreement. Heltrukken linje tilsvare gjennomsnittet av differansen mellom T1 og T2

Deloppgave 2: Vitenskapsfilosofi

Introduksjon

Jeg har valgt å besvare oppgave 2 og 3 i dette arbeidskravet og deler arbeidskravet inn i 2 deler.

Del 1 “Gi en kort beskrivelse og forklaring av falsifikasjonisme. Presenter så ett problem med denne teorien. Tror du problemet kan løses? Begrunn svaret ditt.”

Del 2 “Oppgave 2: Forklar de grunnleggende ideene i Bayesianisme. Hva er noen av fordelene med Bayesianisme sammenlignet med alternativene? (for eksempel, HD-teorien eller abduktivisme). Presenter ett problem med Bayesianisme og evaluer hvor alvorlig problemet er.”

Del 1

Beskrivelse og forklaring av falsifikasjonisme

Falsifikasjon er en teori innen vitenskapsfilosofi som kjennetegnes ved at teorier (hypoteser) er falsifiserbare (motvisbare). Det motsatte er at teorier er verifiserbare (bekreftet).

Gjennom store deler av det 20.århundre var problemet med bekreftelse av teorier en stor utfordring for filosofene.

Gruppen teoretiske empirister bygde opp “induktiv logikk” for å hjelpe oss med å forstå testing i vitenskap. De lagde seg enkle hypoteser, men dette feilet (f.eks. at alle ravner er svarte). (Bayesianism and modern theories of evidence).

Karl Popper likte denne situasjonen da han mente at induktiv logikk var en fiasko fordi induksjonsproblemet er vanskelig/umulig å løse. Derfor sa Popper at “S er en vitenskapelig påstand hvis og bare hvis S er falsifiserbar” (Popper,1959).

Poppers defisjon på falsibarhet er “S er falsifiserbar hvis og bare hvis man kan deduktivt utlede en prediksjon fra S som kan empirisk sjekkes”.

Popper kom med ideen om at vitenskapen måtte “gi slipp på at en teori kan bekreftes” fordi ingen vitenskapelige teorier kan bekreftes. Popper mente videre at induksjonen var vanskelig på grunn av utfordringen med på å skille vitenskapelige hypoteser fra ikke-vitenskapelige hypoteser og at vitenskap kun må bruke falsifisering. Han mente at det bare kunne være falsifiserbare hypoteser for at det skal være ekte vitenskap. Desto mer man kunne gradere falsifiseringen med flere eller mer presise prediksjoner (antagelser), desto bedre vitenskap.

Poppers favoritt teori er Einstein relativitets teori. Den er svært falsifiserbar, og den har overlevd nye prediksjoner og falsifiseringer. Den er det som kalles korroborert (overlevd mange tester), og den er sannsynligvis sann (men ikke bekreftet).

Popper vitenskapelige syklusen; komme opp med nye prediksjoner og falsifiseringer, “så tvil hele veien”, “man kan aldri bli sikker”

Problemet med falsifikasjonisme, Poppers og Duhems læsning på problemet

Innen falsifikasjonisme er en hypotese vitenskapelig hvis, og bare hvis, den er falsifiserbar. Pierre Duhem, en fransk filosof, kom med noen logiske poeng, og utfordret Popper.

Et eksperiment som støter på et problem, vil ikke kunne gi svar på hvor feilen ligger uten støttehypoteser og bekreftelser. Popper ønsket ikke bekreftelser i vitenskap, så her fikk falsi-

fikasjonen et problem. Duhem mente at falsifisering krever (vanligvis) bekreftelse, og bekreftelse krever induksjon!

Avslutter del 1 med Duhems ord: “Det eneste eksperimentet lærer oss er at blant forslagene som brukes til å forutsi fenomenet og for å fastslå om det ville bli produsert, er det minst en feil; men hvor denne feilen ligger, er akkurat det eksperimentet ikke forteller oss.”

Del 2

Bayesianisme og Bayes Theorem

Bayesianisme bruker kvantitativ sannsynlighetsregning og induktiv logikk. Det vil si at man ønsker å forstå bevis ved hjelp av sannsynlighetsteori, som å regne ut sannsynligheten for at noe inntreffer gitt en annen hendelse.

Om det er en usikkerhet omkring en hypotese, vil observasjonsbevis noen ganger heve eller senke sannsynligheten for hypotesen.

Vi snakker ofte om muligheter som mer eller mindre sannsynlige. Forskere snakker også om hypoteser som sannsynlige eller usannsynlige gitt vitenskapelig kunnskap og eksperimentelle funn. Dette er ofte kjent som $P(X)$ er sannsynligheten for at påstanden X er sann. Et enkelt eksempel er: $P(\text{det regner senere})$ er sannsynligheten for at det regner senere.

Thomas Bayes, en engelsk statistiker og matematiker, prøvde seg på et utsagn som viste seg å stemme for å regne ut sannsynlighet for en påstand hvor nye bevis kommer frem og stiller spørsmålstegn ved hypotesen. Dette ble kalt Bayes Theorem. Med Bayes Theorem ønsker vi en rasjonell oppdatering av sannsynligheter ettersom bevis kommer inn.

Formelen for Bayes' teorem er: $P(h|e) = [P(e|h) * P(h)] / P(e)$

Vi leser “h” som hypotese og “e” som bevis. Videre forståelse er:

$P(h|e)$ representerer sannsynligheten for at hendelse h skal skje gitt at hendelse “ e ” har skjedd.

$P(e|h)$ er sannsynligheten for at hendelse e skal skje gitt at hendelse h har skjedd. Sannsynlighet for formen $P(e|h)$ blir ofte kalt “likelihoods” (“sannsynligheter”).

$P(h)$ er den tidligere sannsynligheten for at hendelse “ h ” skal skje uten å vurdere bevis.

$P(e)$ er den tidligere sannsynligheten for at hendelse “ e ” skal skje uten å vurdere bevis.

Bayes teorem forteller oss altså om relasjonen mellom $P(h)$ og $P(h|e)$ og at ved nye bevis “ e ” vil “dagens sannsynlighet bli morgendagens tidligere sannsynlighet”. En oppdatert versjon kan derfor være: $P_{ny}(h) = P_{gammel}(h|e)$.

HD metoden

Står for den Hypotetisk deduktive metoden. Den er ikke matematisk og bruker kvalitativ metode. En god HD teori blir bekreftet flere ganger.

Trinnene i teorien er fra Hempel på 1950 tallet:

Trinn 1: Formuler en teori eller hypotese – en “utdannet gjetning”.

Trinn 2: Utled empiriske konsekvenser fra teorien/hypotesen med deduksjon.

Trinn 3: Test disse empiriske konsekvensene mot observasjoner eller eksperiment.

Trinn 4: Hvis de empiriske konsekvensene viser seg å være riktige, er teorien (induktivt) bekreftet til en viss grad.

Et problem med HD teorien er at den har kun 2 kategorier (falsifisert og bekreftet), og at den ikke skiller mellom enkle og kompliserte hypoteser så lenge de kan bekreftes.

F.eks. dersom 2 teorier som testes kan begge bekreftes (selv om vi egentlig bare ønsker at en av de skal bekreftes). Et prinsipp man bør følge “enkle teorier/hypoteser foretrekkes fremfor vanskeligere teorier” (Ockhams barberkniv). I HD teorien velger man alle data/teori som er bekreftet og skiller ikke på at noen er bedre enn andre.

Adbuktivisme

Denne er også ikke matematisk og den bruker kvalitativ forklaring. Metoden er deduktiv og er også kjent som “slutning til den beste forklaringen” eller at “Ingen annen tilgjengelig teori ville gi samme/like god forklaring”. En teori bør gi en forklaring på mange/flere typer datasett, kalt forklaringskraft. Vi velger enkle forklaringer fremfor kompliserte, kalt enkelhet. F.eks. Relativitets teorien er en veldig enkel teori (enkelhet), men har presise forklaringer (forklaringskraft).

HD-teori og Abduktivisme vs Bayesinisme

Vitenskapelige teorier har til hensikt å fortelle oss hva vi gjerne vil se. Forskjellen på HD-metoden og Abduktivisme, sammenlignet med Bayesinisme er at de to første bruker deduktiv logikk (de får bekreftelse) og kvalitative metoder. Bayesinisme bruker sannsynlighetsregning, induktiv logikk og en kvantitative metode.

Utfordringen med Bayesianisme og evaluering av hvor alvorlig problemet er

En utfordring med bayesinisme er at hypotesene alltid må tallfestes. Utfordringen kommer gjerne som hva er det egentlig $P(h)$ måler? Hvordan kan vi tolke gammel mot ny $P(h)$?

Så om vi ønsker å bruke Bayes theorem, så trenger vi en tolkning av sannsynlighet som vil tillate oss å snakke om tidligere sannsynlighet. Det er dette bayesianerne har utviklet og det kaller det for subjektivistisk tolkning.

Ideen bak subjektiv bayesinisme er at tro og sannsynlighetsregning er basert på en persons tro på et gitt tidspunkt. Det kommer gjerne inn nye bevis eller observasjoner som setter spørsmålsteget på om påstanden er sann eller usann.

Et problem med bayesinisme kan derfor være at sannsynlighetsregningen er subjektiv og at mangel på enighet kan være utfordrende da nye bevis kommer. En annen utfordring kan være

at det er vanskeligere å velge mellom to teorier når sannsynligheten er basert på en subjektiv måte.

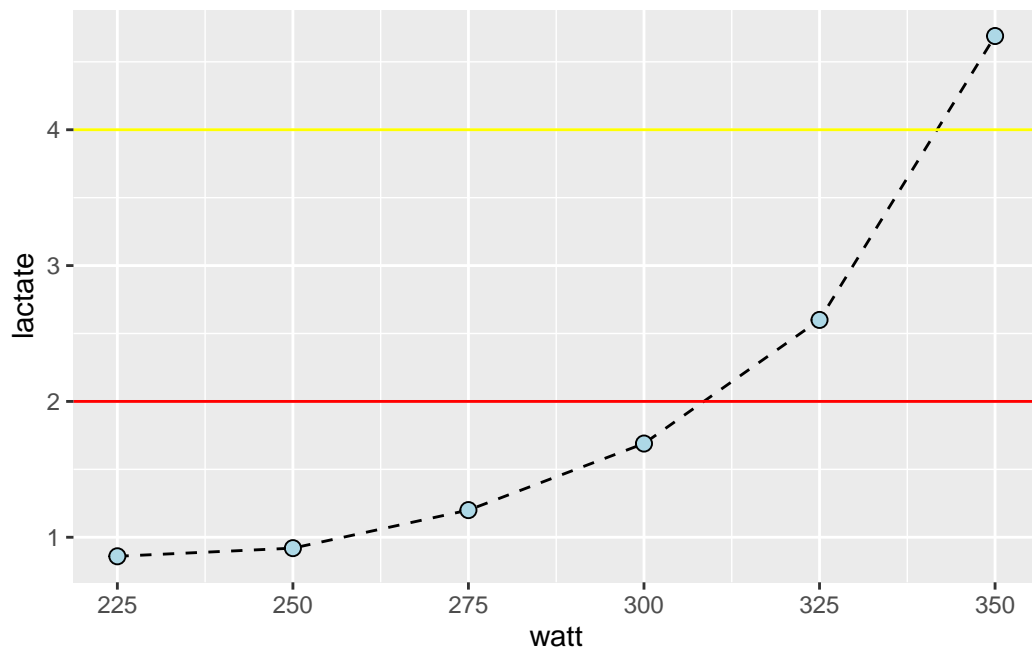
Deloppgave 3: Regresjonsmodeller

Oppgave 1

Undersøke treningsintensiteten på 2mmol L⁻¹ og 4mmol L⁻¹ hos forsøksperson 10 og 3

Her har vi først laget en figur som viser treningsintensiteten på 2 og 4 mmol/L hos forsøksperson 10 og 3. Videre har vi en figur som estimerer laktat terskel etter Machado. Til slutt tall på en laktat terskel etter en anslått fast verdi.

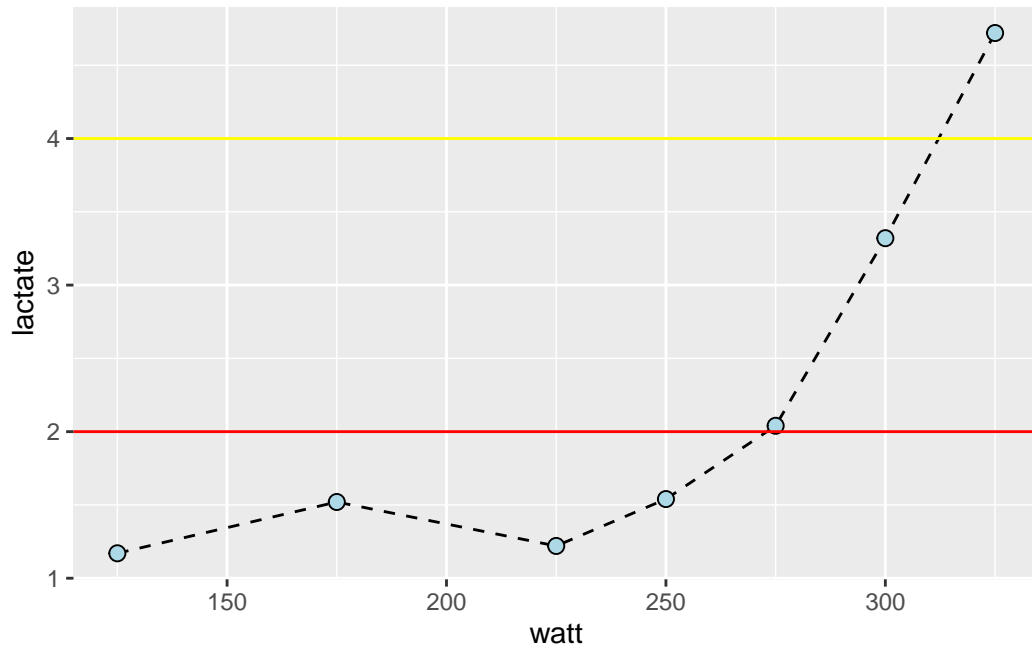
Forsøksperson 10



(Intercept)

3.612157

Forsøksperson 3

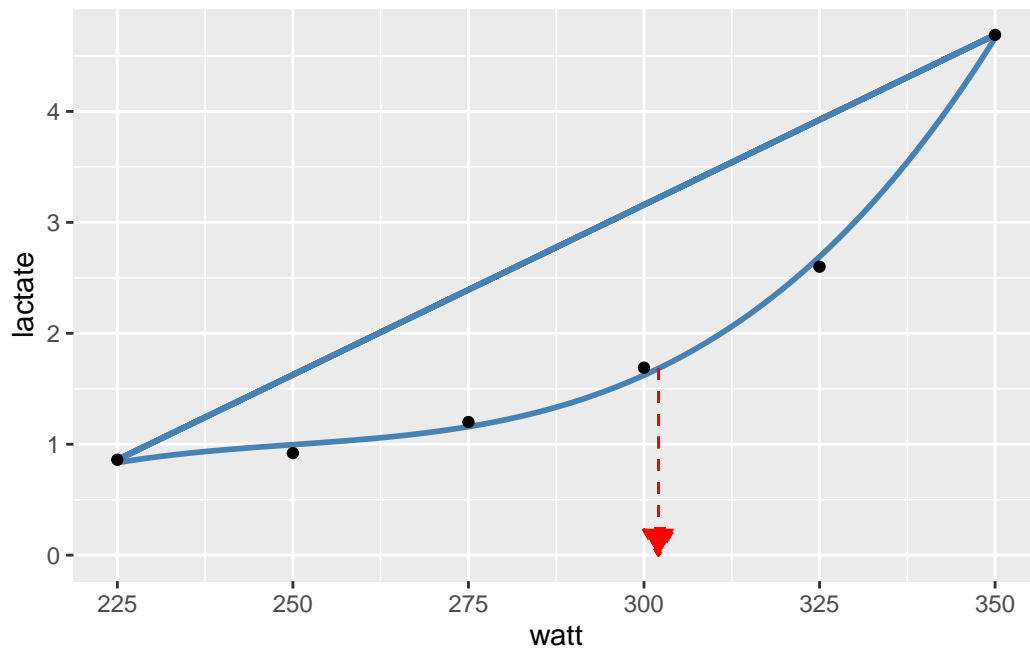


(Intercept)

3.6

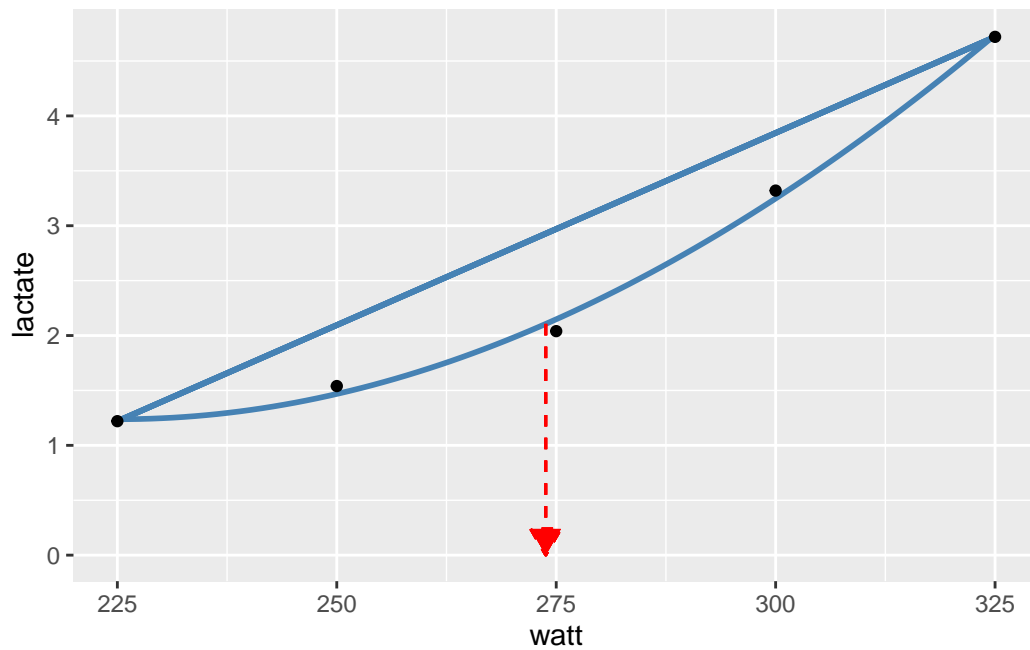
Laktatterskel på forsøksperson 10 etter Machodo “lactate threshold estimated by the maximal deviation method (LTDmax)”.

Forsøksperson 10



Laktatterskel på forsøksperson 3 etter Machado “lactate threshold estimated by the maximal deviation method (LTDmax)”

Forsøksperson 3



“Predicted - fixed lactate value”

Forsøksperson 10

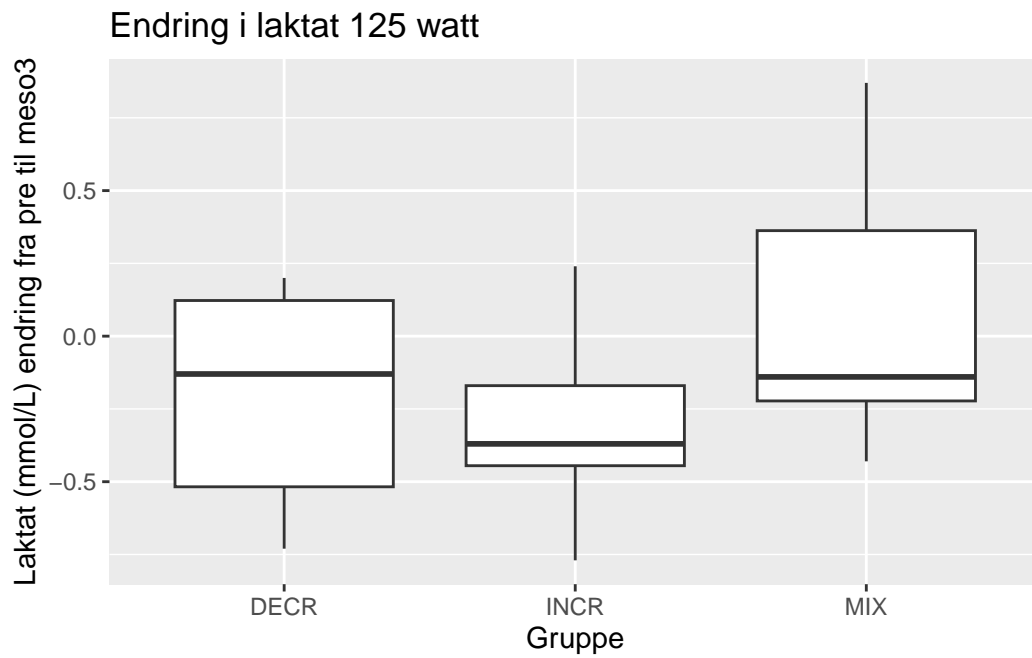
watt	dist
343	0.002613135

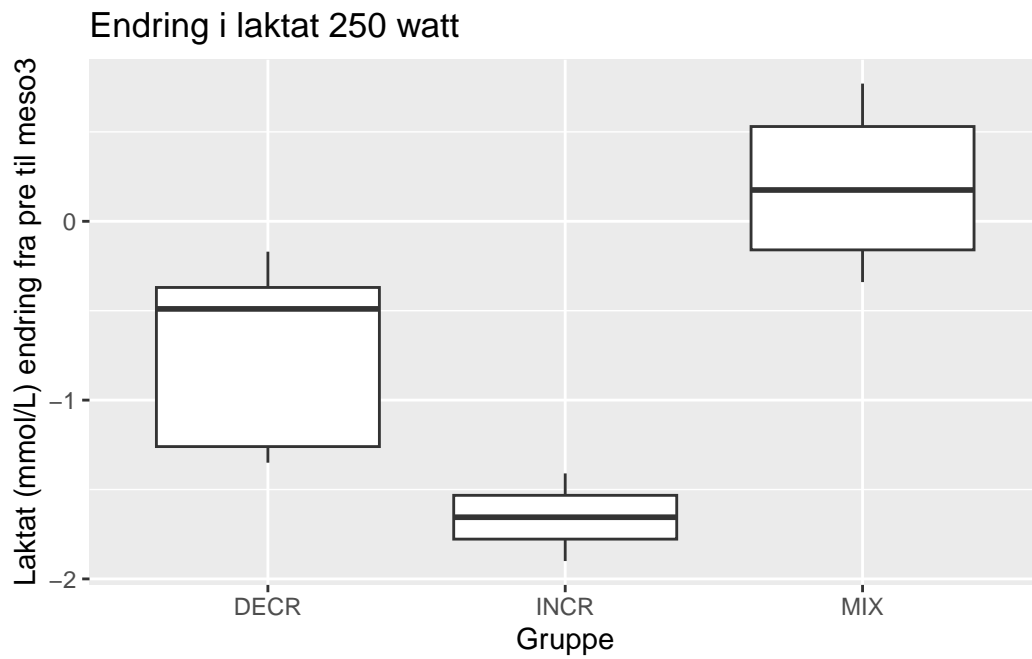
“Predicted - fixed lactate value”

Forsøksperson 3

watt	dist
314.2	0.001172275

Endring i laktat 125 watt og 250 watt

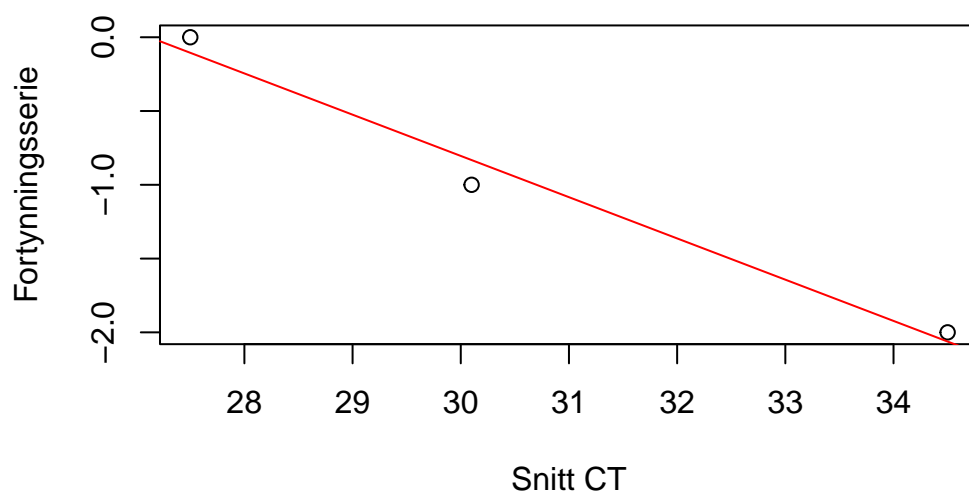




Oppgave 2

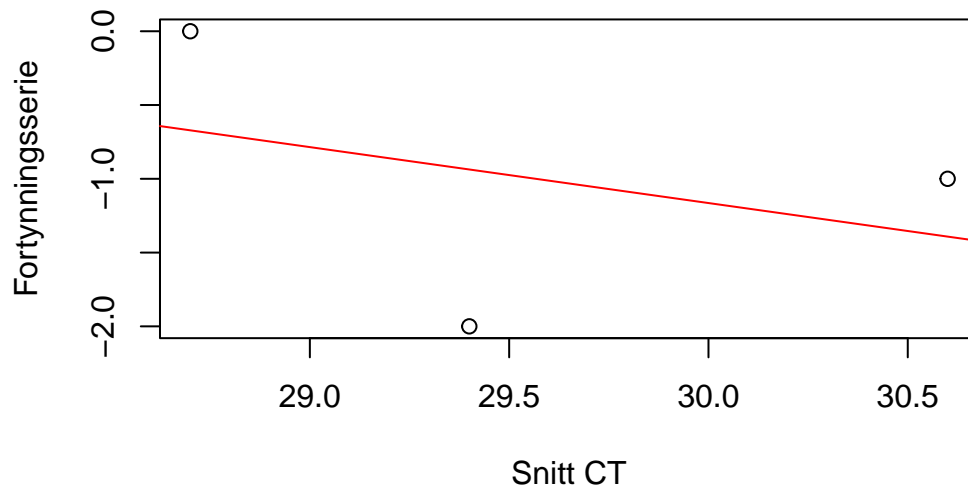
Lage en “calibration curve” som viser effektiviteten av qPCR reaksjonen på “pool sample 1” og “pool sample 2”

Effektivitet av pcr reaksjon 1



Fortynningsserie	Snitt CT	Log
1/1	27.45970	0
1/10	30.07881	-1
1/100	34.50788	-2
1/1000	NA	-3
1/100000	NA	-4
1/100000	NA	-5

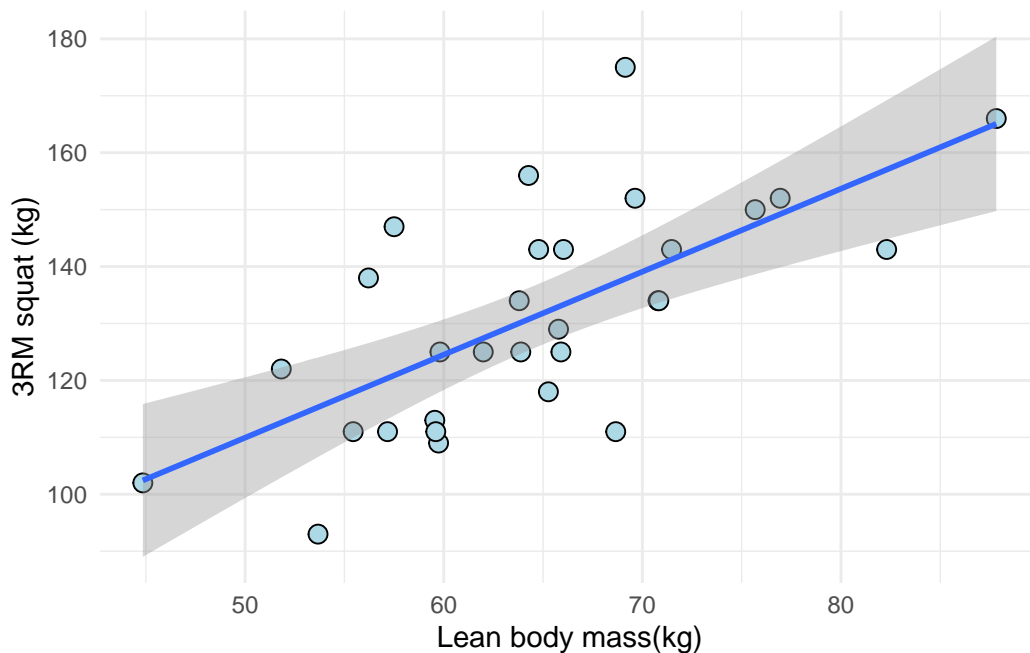
Effektivitet av pcr reaksjon 2



Fortynningsserie	Snitt CT	Log
1/1	28.69725	0
1/10	30.60317	-1
1/100	29.44557	-2
1/1000	NA	-3
1/100000	NA	-4
1/100000	NA	-5

Oppgave 3

Vi skal undersøke om det er en sammenheng mellom økt mager kroppsmasse (LBM) og 3 repetisjoner maks i knebøy (3RM squat).



term	estimate	std.error	statistic	p.value
(Intercept)	37.048	19.705	1.88	0.071
DXA_LBM_T1	1.458	0.302	4.829	4.421×10^{-5}

I tabellen over ser vi sammenhengen mellom lean body mass og 3RM i knebøy.

Estimate: Dette viser stigningstallet. Vi ser at for hver kilogram LBM øker, vil 3RM i knebøy øke med 1,458kg.

Std.error: Standardfeil er et mål for hvor nøyaktig gjennomsnittet av en statistisk distribusjon representerer den sanne verdien. Jo lavere tallet er, jo mer pålitelig er gjennomsnittsestimatet. Tallet på 0,302 viser til en relativt lav spredning i gjennomsnittet.

Statistics: T-verdien er: $\frac{\text{stigningstallet}}{\text{standardfeil}}$. T-verdien brukes til å vurdere signifikansen til hver koeffisient i regresjonsmodellen. Jo høyere t-verdien er, jo større sjanse er det for at det er en signifikant sammenheng mellom LBM og 3RM knebøy. en T-verdi på 4,8 viser til en svært liten sjansje for at sammenhengen er tilfeldig.

P-verdi: Viser til hvor stor grad vi kan anta at det er en signifikant sammenheng mellom den uavhengige og avhengige variabelen. Hvis p-verdien er under 0,05 regnes det ofte som en signifikant sammenheng. Lavere P-verdi viser til lavere sjanse for at sammenhengen vi ser er tilfeldig. vær P-verdi på: $4.421 \cdot 10^{-5}$ viser til en svært stor sammenheng mellom økning i LBM og 3RM knebøy.

Deloppgave 4: Statistisk interferens

ARBEIDSKRAV 3

Call:

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

Residuals:

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

Coefficients:

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

Residual standard error: 3.539 on 7 degrees of freedom

Call:

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

Residuals:

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

Coefficients:

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

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

Residual standard error: 3.019 on 39 degrees of freedom

Call:

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

Coefficients:

(Intercept)	1.84
-------------	------

[1] 1.839727

(Intercept)

1.839727

[1] 3.539191

[1] 1.251293

Explain the estimate, SE, t-value, and p-value from the regression models that we created previously (m1 and m2)

En lineær regresjon har funksjonen $y = ax + b$ og består av et stigningstall/slope (a) og et konstantledd/intercept (b). Det er ikke alltid det er like åpenlyst hvordan likningen skal se ut,

og vi er derfor nødt til å bruke den kunnskapen vi har til å estimere både slope og intercept. Dette gjøres ved at R genererer en likning som avviker minst mulig fra alle datapunktene. Da får vi et estimat til koeffisientene som forteller oss hvor langt unna virkeligheten den estimerte likningen er. I m1 er estimatet 1,290 og i m2 er det 1,4799.

Standard error (SE) forteller oss noe om hvor usikker koeffisienten er, og brukes ofte til å lage konfidensintervall. SE regnes ut slik:

$$SE = \frac{SD}{\sqrt{n}}$$

I m1 er SE 1,290 mens den er 1,480 i m2. Det er veldig like SE men m1 kan se ut til å være noe mer sikker enn m2 da SE er lavere.

T-verdien er koeffisienten delt på SE. Vi vil gjerne at t-verdien skal være høy fordi det betyr at SE er lite sammenliknet med koeffisienten og vi kan være mer sikker på resultatet vårt. I m1 er t-verdien 0,94 og i m2 er den 3,953. T-verdien er videre brukt til å finne p-verdien.

P-verdien forteller oss hvorvidt noe er statistisk signifikant eller ikke. Grensen er ofte satt til 0,05 eller 5%, og dersom p-verdien er lavere enn dette er det statistisk signifikant. Det p-verdien forteller oss er hvor sannsynlig det er at vi vil observere det vi observerte eller noe mer ekstremt dersom vi repeter prosessen igjen. I m1 er p-verdien 0,378 og i m2 er den 0,000315. Her er begge under 0,05 men m2 er tilnærmet lik null.

Discuss what contributes to the different results in the two studies (m1 and m2)

Totalt har vi en populasjon på 100000 stk med individer som har vært gjennom to ulike behandlinger. Ut fra denne populasjonen velger vi ut to grupper. Den ene gruppen (m1) har 8 individer mens den andre gruppen (m2) har 40. Til tross for at de to gruppene består av tilfeldige individer fra den originale populasjonen, kan man ikke være sikker på at så små grupper representerer en hel populasjon da fordelingen i gruppene ikke nødvendigvis tilsvarer fordelingen i populasjonen til tross for randomisering. Dette gjelder særlig m1 da hvert enkelt individ utgjør en mye større prosentandel av gruppen, som gir rom for større skjevheter enn

hva som er tilfellet i den faktiske populasjonen. Et lite utvalg slik som i m1 vil også bety større grad av usikkerhet, som kommer frem i estimatet og SE.

Why do we use the shaded area in the lower and upper tail of the t-distribution

Målet med statistikken er å si noe om en populasjon vha. et utvalg. Hvis utvalget er et tilfeldig utvalg fra populasjonen, har vi mulighet til å estimere egenskaper i populasjonen med en estimert usikkerhet. Figuren viser en normalfordelingskurve. Vi kan beregne hvor mange teoretiske utvalg som vil være mer ekstreme enn f.eks. 95% av alle utvalg, og det er dette som er de mørke områdene i figuren. Med gjentatte utvalg, vil 95% av alle konfidensintervall inneholde gjennomsnittet i populasjonen. De siste 5% er de mest ekstreme tilfellene. Slike figurer illustrerer bl.a. standardavvik og hvor grensen for statistisk signifikans går, som er de mest ekstreme 5%.

Using the “results” data frame, calculate the standard deviation of the estimate variable, and the average of the se variable for each of the study sample sizes (8 and 40). Explain why these numbers are very similar. How can you define the Standard Error (SE) in light of these calculations?

```
[1] 0.4696954
```

```
[1] 1.021374
```

```
[1] 0.4838475
```

```
[1] 1.070843
```

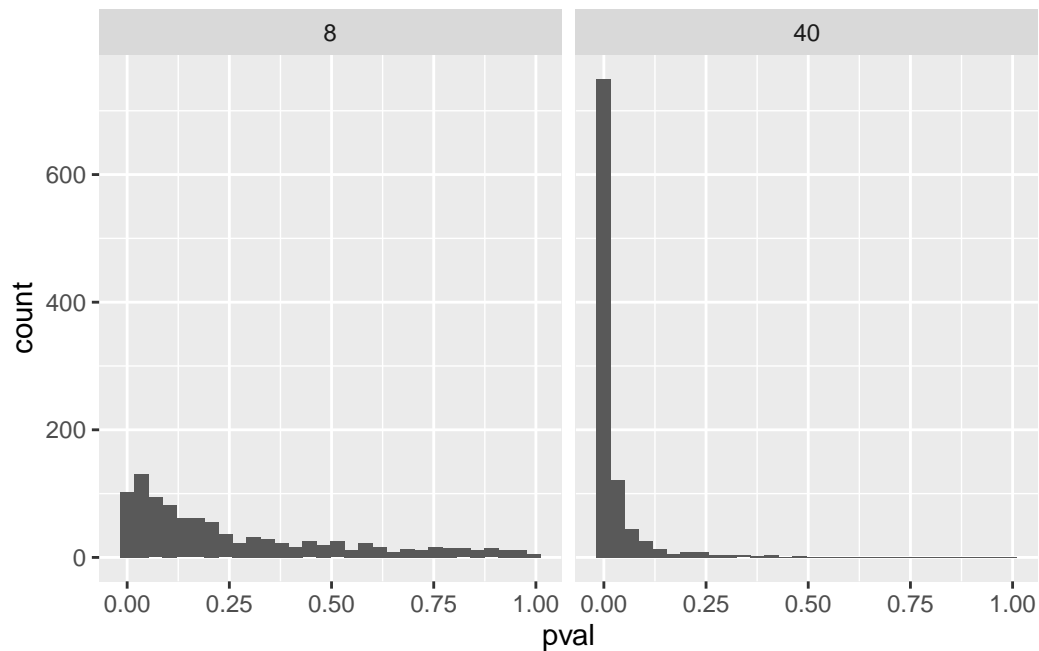
Gjennomsnittet til “se” i gruppen med 40 er ~0,470 eller 47% mens det i gruppen med 8 er ~1,021 eller 102%. Det er et stort standard error i begge gruppene, men et meget stort standard error i gruppen med kun 8 stk. Dette viser til hvor stor usikkerhet det er i “se” resultatene fra dataene. For eksempel kan resultatene i gruppen med 40 variere med 47% både opp og ned.

Standardavviket til “estimate” i gruppen med 40 deltakere er ~0,484 mens det i gruppen med 8 er ~1,071. Standardavviket forteller oss om spredningen av dataene rundt gjennomsnittet. Et lavt standardavvik betyr at det er lite spredning mens et større standardavvik betyr en større spredning som betyr større usikkerhet.

For både “estimate” og “se” ser vi at usikkerheten er mye større i gruppen med kun 8 sammenliknet med gruppen med 40. Dette bidrar til å vise hvor viktig utvalgsstørrelsen er for resultatene og med hvor stor sikkerhet vi kan stole på dataene.

Using the “results” data frame, create a histogram (see example code below) of the p-values from each study sample-size. How do you interpret these histograms, what do they tell you about the effect of sample size on statistical power?

``stat_bin()` using `bins = 30`. Pick better value with `binwidth`.`



```
# A tibble: 2 x 2
  n sig_results
```

	<dbl>	<dbl>
1	8	0.227
2	40	0.865

One-sample t test power calculation

```

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

```

Using the “results” data frame, calculate the number of studies from each sample size that declare a statistical significant effect (specify a threshold for “a” your significance level)

n	antall
8	227
40	865

I datasettet med 40 er det 865 som har en statistisk signifikant effekt, mens det i datasettet med 8 kun er 227 som har en statistisk signifikant effekt.

Using the pwr package, calculate the power of a one-sample t-test, with a effect size of 1.5/3, your specified significance level and sample sizes 8 and 40. Explain the results in the light of your simulations

One-sample t test power calculation

```

n = 40

```



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

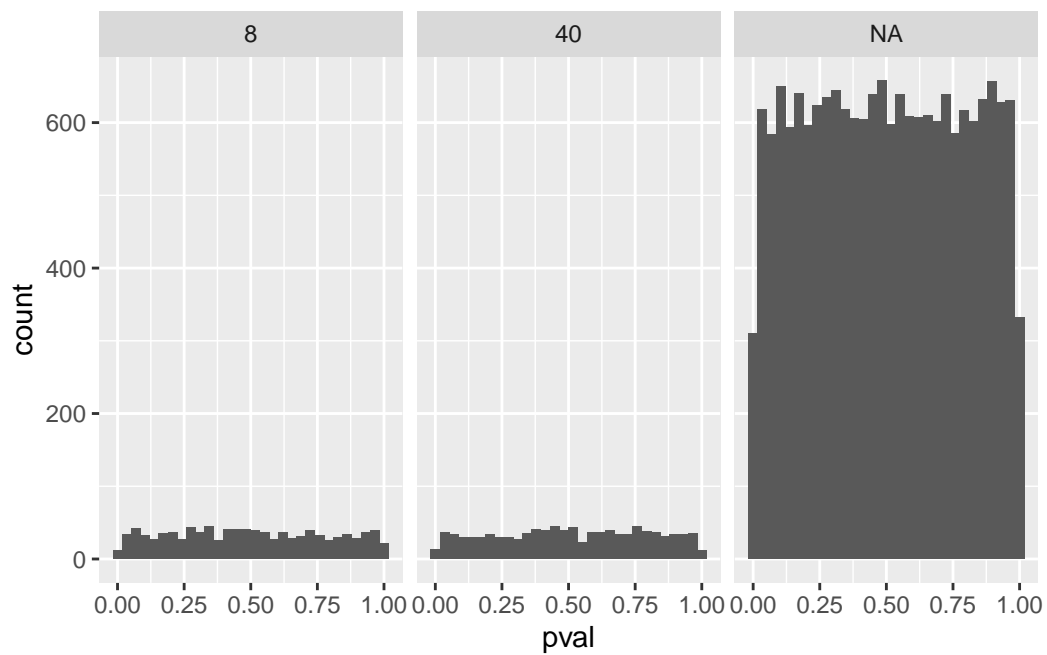
One-sample t test power calculation

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

For gruppene med både 40 og 8 er effektstørrelsen (d) 0,5, mens “power” er ~0,869 i gruppen med 40 og ~0,232 i gruppen med 8. Testens “power” er et tall mellom 0 og 1 som indikerer hvor sannsynlig det er å observere en signifikant endring dersom den faktisk eksisterer en endring. Jo nærmere 1 “power” er, desto mer sannsynlig er det at testen klarer å påvise en endring. Her kommer et også frem at det er en stor forskjell mellom gruppen med 8 og gruppen med 40, der gruppen med 40 har en mye høyere power enn gruppen med 8.

Using the new data frame with results from studies of a population with an average effect of zero, create new histograms with a significance level of 5%. How many studies would give you a “false positive” result if you did many repeated studies?

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Dersom vi har 2000 observasjoner, slik som i `results_null`, og signifikansnivået settes til 0,05 eller 5% kan vi anta at 100 av disse vil gi en falsk positiv. $0.05 * 2000 = 100$

Deloppgave 5: Studiedesign

Styrketrening for eldre

Introduksjon

Jeg har valgt ut å se på effektene av styrketrening hos eldre. Et litteratur søk via skolens google scholar og bibliotek ble gjort. Det er ulike studie design og de jeg endte opp med å bruke i denne oppgaven har ulike kvalitet for bevis. Meta-analyser, Randomiserte kontrollerte studier, case kontroll, case-kontroll studier og kasus-rapporter. Figur 1 i (ahnIntroductionSystematicReview2018?) viser de ulike designene. På toppen av ligger de systematisk oversiktene og meta-analysene som samler alle mulige studier knyttet til et gitt tema og design, og gjennomgår og analyserer resultatene deres. En metaanalyse er en gyldig, objektiv og vitenskapelig metode for å analysere og kombinere ulike resultater. Vanligvis, for å oppnå mer pålitelige resultater, utføres en metaanalyse hovedsakelig på randomiserte kontrollerte studier (RCT), som har et høyt bevisnivå. Videre har jeg sett på kasus-kontroll studier, kasus studier og kasus rapporter.

Alle artiklene ser på styrketrening for eldre, men med litt ulike effekter da noen besvarer ett spørsmål, mens andre besvarer flere spørsmål. Jeg har også prøvd å se på protollene for styrketrening som er beskrevet.

I alle studiene er det en god introduksjon som gjør at man kjenner seg igjen i hva som skal undersøkes. Jeg har prøvd å lage en sammenfatning av det viktigste i introduksjonene til de ulike artiklene.

Ifølge data fra FN vil verdens befolkning over 65 år øke fra 9.7% i 2022 til 16.4% i 2050. Fall hos eldre regnes som et av de alvorligste og mest kostbare folkehelseproblemene i når det gjelder sykkelighet, dødelighet og enorme kostnader for helse- og sosialtjenester.

Fall hos eldre er en utfordring og deles inn i ytre og indre faktorer. De ytre årsakene er vanligvis relatert til hjemmemiljøet som blant annet tepper, trapper og dårlig belysning. De indre faktorene er relatert til reduksjon i balanse- eller gangfunksjon, muskelsvakhhet, nedsatt syn, nedsatt kognitive funksjon, sykdommer og bivirkninger av medisiner. Studier har vist at de mest effektive intervensjonene for å forebygge fall er basert på balansetrening og muskelstyrking i underekstremitetene. ((ishigakiEffectivenessMuscleStrengthening2014?))

Muskelstyrke er vurdert til en svært viktig faktor for fysisk form og livskvalitet. For eldre er styrke trening viktig for å opprettholde funksjonsevne da styrke progressivt reduseres ved økende alder som påvirker evnen til å gjøre lystbetonte hverdagsaktiviteter ((silvaInfluenceStrengthTraining2014?))

I artikkelen til (benekaResistanceTrainingEffects2005?) blir det beskrevet at muskelstyrke reduseres med 1-2% årlig hos personer over 65 år med en høyere forekomst på underekstremitetene versus overekstremitetene. I (ardaliDailyAdjustableProgressive2014?) blir det beskrevet en reduksjon i styrke på 1.5% i alderen 50 til 60 år, og deretter 3% reduksjon i årene etter fylte 60 år.

Spesielt styrke i m.quadriceps har blitt undersøkt og viser progressiv reduksjon i muskelstyrke og det er sett at denne reduksjonen i styrke kan reverseres med øvelser. Forbedring i styrke ved hjelp av styrketrening er sett ved flere ulike typer trening: isokinetisk, isometrisk og isotonisk trening. På studier har man sett en slik treningseffekt med treningsfrekvens på 2-3 ganger pr uke og varighet på 8-24 uker ((benekaResistanceTrainingEffects2005?)).

I følge (ardaliDailyAdjustableProgressive2014?) viser reduksjon i styrke i m.quadriceps en langsommere gang hastighet ved vanlig gange og ved trappegange, samt økende risiko for fall.

Spørsmål:

Studier har gjerne et primær spørsmål man ønsker å undersøke. Samtidig velger flere studier å se på sekundære spørsmål om andre prediktorer eller utfall som også kan gi verdifulle konklusjoner ((**hulleyDesigningClinicalResearch2013a?**)).

Sekundære spørsmål

Det er mange tverrsnittsstudier og kohortstudier ((**hulleyDesigningClinicalResearch2013a?**)) innenfor feltet styrketrening for eldre hvor målet på studiene bl.a. er å se på økende muskelstyrke ved en intervensjon. I (**ishigakiEffectivenessMuscleStrengthening2014?**) har de sett på flere studier og de konkluderer med at den metodologiske kvaliteten på studiene på dette området viser god effekt på fall forebygging hos eldre ved bruk av styrketrening på underekstremitetene. Allikevel er intervensjonene på mange studier dårlig beskrevet.

Design

Metodedelen bør inneholde informasjon om deltagerne, materialer eller apparater, prosedyrer og data-analyse ((**carterRehabilitationResearchPrinciples2011?**)).

(**ardaliDailyAdjustableProgressive2014?**) er en kasusrapport med et godt beskrevet studie design. Den er av typen forskning er plassert nederst i hierarkiet av klinisk bevis ((**ahnIntroductionSystematicReview2018a?**)). Hensikten med denne kasusrapporten er å introdusere og beskrive en tidlig bruk av en protoll for daglig progressiv styrketrening (DAPRE protokoll) som et supplement til standard rehabilitering for å maksimere quadriiceps muskelstyrke og funksjonell ytelse hos en eldre hjemmebundet pasient etter en ny operert totalprotese i kne.

DAPRE er en strukturert progressiv styrketeknikk utviklet klinisk av Knight (1990) i et forsøk på å gi et objektivt middel for å øke motstanden samtidig med styrkeøkninger under knereha-

bilitering etter skade eller operasjon ((**knightQuadricepsStrengtheningDAPRE1990?**)). DAPRE-systemet bruker 4 serier med 5-RM til 7-RM (repetisjon maksimum) med ulike styrkeøvelser. Nøkkelen til DAPRE-teknikken er den konstante tilbakemeldingen som pasienten og terapeuten får under styrkerehabiliteringsøkter. Teknikken tillater dermed individuelle forskjeller i styrkeøkninger og gir en objektiv metode for å øke motstanden i takt med styrkeøkninger ((**knightQuadricepsStrengtheningDAPRE1990?**)).

I meta-analysen til (**csapoEffectsResistanceTraining2016?**) sammenlignet de effekten mellom tung (~80% av 1RM) og lett/moderat styrketrening(~45% av 1 RM). Spørsmålene de ønsket å besvare var om disse ville øke skjelettmuskelmasse (målt ved ultralyd) og styrke (målt ved bl.a. isometrisk leg extension og ved nedtrekk) hos eldre. For å hente artiklene til denne metaanalysen ble det systematisk søkt i PUBMED og MEDLINE. Protokollene for tung styrketrening var gjerne 3 x 8 reps eller 5 x 4 reps. For lett/moderat styrketrening var dette gjerne 3 x 15 eller 3 x 20 reps.

I følge (**hulleyDesigningClinicalResearch2013a?**) kan metaanalyser øke muligheten til å svare på et forskningsspørsmål, men har den ulempen at de ikke inkluderer data på individnivå da de ikke kan individuelt tilpasse belastning som ved kasusrapporter.

Systematiske oversikter og metaanalyser presenterer resultater ved å kombinere og analysere data fra ulike studier utført på lignende forskningstemaer ((**ahnIntroductionSystematicReview2018a?**)).

(**monteiroEffectThreeDifferent2020?**) er en longitudinell randomisert kontrollert studie. Deltakerne gjennomgikk et treningsprogram tre ganger i uken i åtte måneder. Målet var å vurdere effekten av tre forskjellige treningsprotokoller i quadriceps og hamstrings isokinetisk muskelstyrke hos uavhengige eldre kvinner. 80 kvinner ble rekruttert til denne studien og delt inn i fire grupper på tjue personer: kontrollgruppen, multikomponentgruppen, muskelkraftgruppen og muskelstyrkegruppen. Den isokinetiske styrken ble målt i begge underekstremitetene med et dynamometer ved 60°/s og 180°/s.

(**benekaResistanceTrainingEffects2005?**) sin case-kontroll studie hadde to mål. Det første målet med denne studien var å bestemme effekten av styrketreningsintensitet på

styrkeutviklingen hos friske inaktive eldre etter en treningsperiode på 16 uker. Det andre målet var å teste hypotesen om at størrelsen på de intensitetsrelaterte effektene er relatert til leddets vinkelhastighet, og det tredje målet var å undersøke om motstandstrening gir kjønns spesifikk tilpasninger eller ikke.

I (silvaInfluenceStrengthTraining2014?) er en meta-analyse hvor de ønsker å finne den optimale dosen eller mengden styrketrening som vil øke styrken hos personer over 55 år. De undersøkte et dose-responsforhold mellom styrkeforbedring og treningsvariablene; treningsvolum (antall repetisjoner og sett per øvelse), treningsintensitet (% av 1 RM) og treningsvarighet.

De har brukt den standardiserte gjennomsnittlige forskjellen (SMD) som effektstørrelsesindeks til metaanalysen. SMD ble beregnet ved å bruke forskjeller i styrkeforbedring fra baseline mellom forsøkspersoner i styrketreningsgruppen og kontrollgruppen. Effektstørrelsesberegningen var basert på gjennomsnittet,

Det ble brukt RM- tester for kneekstensjon for å undersøke styrkeøkningene. Mange av studiene fulgte American College of Sports Medicine sine anbefalinger for unge voksne (3 sett med 10–15 reps av hver øvelse, frekvens på 2–3 dager i uken). Studiene som ble sett på hadde en variasjon i treningsintensitet på 40-84% av 1RM, men kun 4 av de 15 studiene brukte intensitet under 70% av 1RM.

Tabell 1: Karakteristika i studiene til Silva et al. (2014)

Inkluderte studier	Alder (år)	Fysisk aktivitetsnivå	Kjønn	Statistisk analyse	N	Test
Frontera et al.	68-79	Stillesittende	Kvinne	Anova	14	Kne ekstensjon
Judge et al.	71-97	Stillesittende	Begge	Student t -test	31	Kne ekstensjon
Judge et al.	75-85	Stillesittende	Begge	MANOVA	55	Kne ekstensjon
Kalapotharakos et al.	60-74	Stillesittende	Begge	ANOVA	33	Kne ekstensjon
Morganti et al.	59 ± 0.9	Stillesittende	Kvinne	Student t -test	39	Kne ekstensjon
Pyka et al.	61-78	Stillesittende	Begge	Student t -test	14	Kne ekstensjon
Schlicht et al.	61-87	Stillesittende	Begge	ANCOVA	22	Kne ekstensjon

Seynnes et al.	81.5 ± 1.4	Stillesittende	Begge	Pearsons r	22	Kne ekstensjon
Taffe et al.	65-79	Stillesittende	Begge	ANOVA	46	Kne ekstensjon
Hurley et al.	60 ± 5	Stillesittende	Kvinne	ANOVA	35	Kne ekstensjon
Reeves et al.	74.3 ± 3.5	Stillesittende	Begge	ANOVA	18	Kne ekstensjon
Debeliso et al.	71.6 ± 5.3	Stillesittende	Begge	ANOVA	26	Kne ekstensjon
Rabelo et al.	60-76	Stillesittende	Kvinne	ANOVA	61	Kne ekstensjon
Fatouros et al.	71.2 ± 4.1	Stillesittende	Mann	ANOVA	52	Kne ekstensjon
Debeliso et al.	71.6 ± 5.3	Stillesittende	Begge	ANOVA	60	Kne ekstensjon

Alder er beskrevet som snitt +- SD eller som bredde. N er antall personer i studien.

Statistiske analyser

Tabell 1 viser ulike statistiske tester som ble brukt i de inkluderte studiene til (**silvaInfluenceStrengthTraining2014?**)

Tabellen viser også en fin fremstilling og sammendrag på hva som ble undersøkt og hvilken populasjon dette gjaldt. I (**silvaInfluenceStrengthTraining2014?**) nevner de også effektstørrelse innenfor en meta-regresjon som er hvordan forskjellige faktorer påvirker effektstørrelsen på tvers av ulike studier som inngår i en meta-analyse.

Effektstørrelse er et uttrykk for effekten av en intervensjon som f.eks. en gjennomsnittlig forskjell.

Analyse av variasjon (ANOVA) som er beskrevet i (**carterRehabilitationResearchPrinciples2011?**)

er en parametrisk statistisk test hvor variansen brukes for å sjekke riktigheten. I ANOVA test kan flere variabler testes (f.eks. forskjellen i snitt på flere enn to grupper). ANOVA test er en null hypotese signifikans test utviklet av Ronald Fischer i 1918. En F-verdi brukes for å beregne sannsynlighet. I (**benekaResistanceTrainingEffects2005?**) ble ANOVA og MANOVA testene brukt.

Student t-test er statistisk null hypotese signifikans test hvor kun 2 variabler testes. For eksempel å teste effekten av en ny behandling eller trening og undersøke effekt størrelsen

(**spiegelhalterArtStatisticsLearning2020?**).

Analyse av kovariasjon (ANCOVA) er en statistisk analyse med høy styrke som er beskrevet i (**vickersUsePercentageChange2001a?**). Den kombinerer ANOVA test og lineær regresjonsanalyse. ANCOVA sammenligner endringsscore og estimatene viser de justerte forskjellene mellom gruppene hvis de hadde nøyaktig samme utgangspunkt.

Pearsons korrelasjonskoeffisient måler styrken av den lineære sammenhengen mellom to variabler (**spiegelhalterArtStatisticsLearning2020?**). I kasus rapporten til@ardaliDailyAdjustableProgressive2014 bruker de Pearsons r som mål på effektstørrelse. Pearsons r kan fint brukes i enklere sammenhenger, mens som mål på sammenhenger mellom flere variabler er den ikke egnet.

I andre artikler jeg har sett på som i (**monteiroEffectThreeDifferent2020?**) ble den ikke-parametriske Kolmogorov-Smirnov-testen brukt for å se om en test er signifikant forskjellig fra normalfordelingen og Levene-testen gjorde det mulig å evaluere homoskedastisiteten. De brukte også ANOVA for å vurdere de fire gruppenes forskjeller i de to ulike momentene. Testen gjorde det mulig å evaluere forskjellene i hver gruppe og øyeblikk. Bonferronis korreksjon ble også brukt her.

Resultat

DAPRE-teknikken og funksjonell trening ble funnet å være effektive til å forbedre quadriceps-forlengelsesstyrken og funksjonell ytelse som gange, balanse og trapper hos denne pasienten etter TKA over fem uker(**ardaliDailyAdjustableProgressive2014?**)).

Ved å oppsummering av resultatene til (**csapoEffectsResistanceTraining2016?**) fant de at treningsprogrammer med både høy og lettere belastning indusere betydelig styrkeøkning sammenlignet med ikke-treningskontroller. På spørsmål 2 så de en gjennomsnittlig økning i muskelstørrelse på 11% på gruppen som trente tung styrketrening og 9% med gruppen med lett/moderat styrketrening.

Styrketreningen med 3-4 sett med 3-6 repetisjoner (40-60% av 1RM) med 3-5 min pause ser ut til å være den mest effektive for å øke den isokinetiske styrken hos eldre uavhengige kvinner. Alle treningsprogrammene ser ut til å bidra til å redusere det aldersrelaterte tapet på isokinetisk styrke (**monteiroEffectThreeDifferent2020?**).

HI (90% av 1 RM) -treningsprotokoll viste ifølge (**ardaliDailyAdjustableProgressive2014?**) den beste forbedringen av kneekstensorens isokinetiske kraft sammenlignet med LI (50% av 1 RM) eller MI (70% av 1 RM) ved alle testhastigheter. Imidlertid hadde også treningsgruppene med moderat og lav intensitet også en betydelig styrkeforbedring.

Tilslutt i resultatdelen ser vi at metaanalysen til (**silvaInfluenceStrengthTraining2014?**) viser at flere ulike treningsvariabler (intensitet, antall sett per muskelgruppe, ukentlig treningsfrekvens) kan være like effektive for å forbedre styrken til friske stillesittende eldre mennesker. Den eneste treningsvariabelen som viste en signifikant effektstørrelse for styrkeøkning, innenfor en meta-regresjon inkludert de andre variablene, var treningsvarighet. Dette funnet indikerer at jo lengre varighet av styrketreningsprogrammer, desto større styrkeøkning oppnår eldre individer, uavhengig av ulike kombinasjoner av treningsintensitet, frekvens eller antall sett.

Konklusjon

Alle studiene jeg har sett på viser god effekt på muskelstyrkeforbedring ved bruk av styrketrening i denne aldergruppen. Allikevel viser det seg at høyere intensitet på styrketreningen gir bedre effekt enn moderat og lett intensitet. Det viser seg også at varigheten/treningsvolumet på styrketreningen er en viktig faktor for muskelstyrkeforbedring hos eldre. Konklusjonen i metaanalysen til (**silvaInfluenceStrengthTraining2014?**) er på høyeste nivå av bevis (**ahnIntroductionSystematicReview2018?**) og dette gir informasjon om hva som er god styrketrening hos eldre. Videre forskning på feltet trengs, f.eks. design på hvordan andre faktorer påvirker responsen på styrketrening. Som f.eks. sykdom eller skader, da dette er normale faktorer ved økende alder.

References