

Mappeeksamen IDR4000

Ole-Kristian Kjølner

2024-09-10

Table of contents

Introduksjon	4
1 Assignment 1: Reliability and tools for reproducible data science	5
1.1 Introduksjon	5
1.2 Metode - gjennomføring av $V O_{2max}$ test	5
1.2.1 Utstyr som ble benyttet	5
1.2.2 Forberedelse og kalibrering av utstyr	6
1.2.3 Forberedelse av testdeltager	6
1.2.4 Når testdeltaker er ferdig å sykle	7
1.3 Standardisering av test	7
1.4 Intern validering	8
1.5 Tabeller og plots med datasett fra alle gruppene	8
1.5.1 Plot	8
1.5.2 Tabell	8
1.5.3 Beregning av reliabilitet	9
2 Assignment 2: Regression models, predicting from data	10
2.1 Introduksjon	10
2.2 Part 1 - Lactate thresholds	10
2.2.1 Metode	10
2.2.2 Resultat	10
2.2.3 Diskusjon	12
2.3 Part 2 - Predicting sizes of DNA fragments	13
2.3.1 Metode	13
2.3.2 Resultat	13
2.3.3 Diskusjon	15
2.4 Part 3 - Interpreting a regression table	15
2.4.1 Metode	15
2.4.2 Resultat	15
2.4.3 Diskusjon	16
2.5 Referanser	17
3 Assignment 3: Drawing inference from statistical models, and statistical power	18
4 Assignment 4: Study designs	19
4.1 Overview	19

5	Assignment 5: Analyzing repeated measures experiments	20
5.1	Assignment overview	20
5.2	Introduction	20
5.3	Methods	20
5.3.1	Participants and study overview	20
5.3.2	Muscle strength and hypertrophy	20
5.3.3	Data analysis and statistics	20
5.4	Results	20
5.5	Discussion	24
5.6	Conclusion	24
6	Philosophy of science	25
7	Molecular Laboratory report	26
	References	27

Introduksjon

Mappeeksamen består av følgende deler:

- Rapport: “Deskriptiv statistikk, reliabilitet og validitet og verktøy for reproduserbar vitenskap”.
- Laborasjonsrapport fra molekylærlabb
- Arbeidskrav i vitenskapsteori
- Rapport: “Statistisk inferens, statistiske modeller og statistisk styrke”
- Rapport: “Studiedesign”
- Rapport: “Analyse av eksperimenter med repeterte målinger”

1 Assignment 1: Reliability and tools for reproducible data science

1.1 Introduksjon

1.2 Metode - gjennomføring av $\dot{V} O_{2\max}$ test

1.2.1 Utstyr som ble benyttet

- Biosen (for måling av laktat)
 - lansett finger
 - rør for oppsamling av blod
 - blandingscontainer for rør med blod
 - papir for å tørke av finger
- Vyntus (for måling av metabolsk respons)
 - miksekammer
 - turbin
 - slange fra miksekammer til munnstykke
 - neseklype (teipbit på nese for at neseklype skulle sitte godt)
 - gassbeholder med referanse-gass
- Lode-sykkel
 - Samme oppsett på sykkel ved hver test (høyde sete, avtsand sete til styre, osv.). Oppsettet lagres på data etter innstilling ved t1.
 - MTB pedaler
 - 172,5 mm lengde krankarmer
- Gulvvifte (samme innstilling ved hver test)
- Pulsbelte garmin
- Hansker
- Minnepenn for overføring av data

1.2.2 Forberedelse og kalibrering av utstyr

1. Sørg for at Biosen er på, og at det er nok væske i kalibreringssolution (rød flaske). Sett igang kalibrering av Biosen om det er nødvendig.
2. Ta med nødvendig laktatutstyr bort til sykkelen, herunder: lansett, rør for oppsamling av blod og papir.
3. Kalibrering av Vyntus CPX
 - a. Skru på kalibreringsgassen.
 - b. Se til at turbin er koblet til sampling slange og optoelektronisk måler.
 - c. Koble til kalibreringsporten på CPX-en og kjør både volum- og gasskalibrering. OBS! Ved differanse på over 2% under volumkalibrering eller differanse over 0.2% differanse under gasskalibrering, gjennomfør nødvendig feilsøking og kjør kalibrering på nytt.
4. Ta på hansker og sett sammen munnstykke. Dekk til munnstykket med papir med hjelp av neseklype.
5. Ta med munnstykke og slange bort til sykkel og Vyntus CPX. Fest munnstykke til slange, og slange til miksekammer.
6. Legg til ny eller velg aktuell testdeltager.
 - a. Ny testdeltager legges til, og følgende ID-informasjon legges til under fornavn, etternavn og personnummer: idr4000_h24_gr4_idxx (xx=nr).

1.2.3 Forberedelse av testdeltager

- Informasjon fra testansvarlig til testdeltaker vedrørende hvordan test skal gjennomføres:
 - VO₂max test der vi starter på en gitt watt (motstand). Watten vil øke hvert minutt (hvor mye avhenger av protokoll) og testdeltaker sykler helt til han/hun ikke greier mer.
 - Test avsluttes dersom tråkkfrekvens < 60 rpm.
 - Test skal gjennomføres sittende
 - Borg skala benyttes når utmattelse inntreffer hos testdeltaker. Deltaker vil bli spurt om å gi et tall fra 6-20 på hvor sliten han/hun er, der 6 er “ingen anstrengelse” og 20 er “maksimal anstrengelse”.
 - Underveis i test vil testdeltaker ha tilgang på hvor lenge han/hun har syklet.
 - Testdeltaker informeres om at testleder kun vil informere om hvor lenge det er til neste økning og hvilken watt som sykles på for øyeblikket. Annen data som puls, VO₂, osv., vil ikke utøver ha tilgang på underveis i test.
 - Testleder vil mot slutten av test bidra til å pushe testdeltaker slik at han/hun får ut sitt ytterste. Hvordan denne “pushingen” gjennomføres vil variere fra testleder

til testleder. Det viktige er at testleder oppfører seg tilnærmet likt ved hver test ovenfor den aktuelle testdeltaker.

1.2.4 Når testdeltaker er ferdig å sykle

- Testleder har nå spurt om Borgs skala
- Etter 1 min tas en laktatprøve av testdeltaker for å estimere la_{max} . Prøver plasseres i beger med løsning før den vendes og plasseres i Biosen for måling.
- Noterer ned hvor lenge deltaker syklet og henter så ut resten av data fra test via rapport på vyntus.
- Data som er hentet ut er formatert i excel. Deretter benyttet vi `read_xlsx` for å importere datafilen til R.

1.3 Standardisering av test

- Matinntak:
 - siste store måltid senest 2 timer før teststart. Tillatt å spise en mindre karbohydratkilde senest 30 min før test, dvs. en banan, en bar, en gel, osv.
- Koffeininntak:
 - inntak som normalt.
- Test gjennomføres på samme tidspunkt på dagen ved alle tester
- Søvn/døgnrytme:
 - forsøk å opprettholde normal døgnrytme gjennom testperioden.
- Trening:
 - ingen hard trening på underkropp dagen før test.
- Oppvarmingsprotokoll:
 - 5 min oppvarming, der deltageren sykler på økende intensitet i 2-2-1 min, eks: (2 min 150W, 2 min 175W, 1 min 220W)
- Samme testprotokoll ved alle testene:
 - Her vil det være individuelle forskjeller, eks: start på 200W - 20W økning hvert minutt til utmattelse.

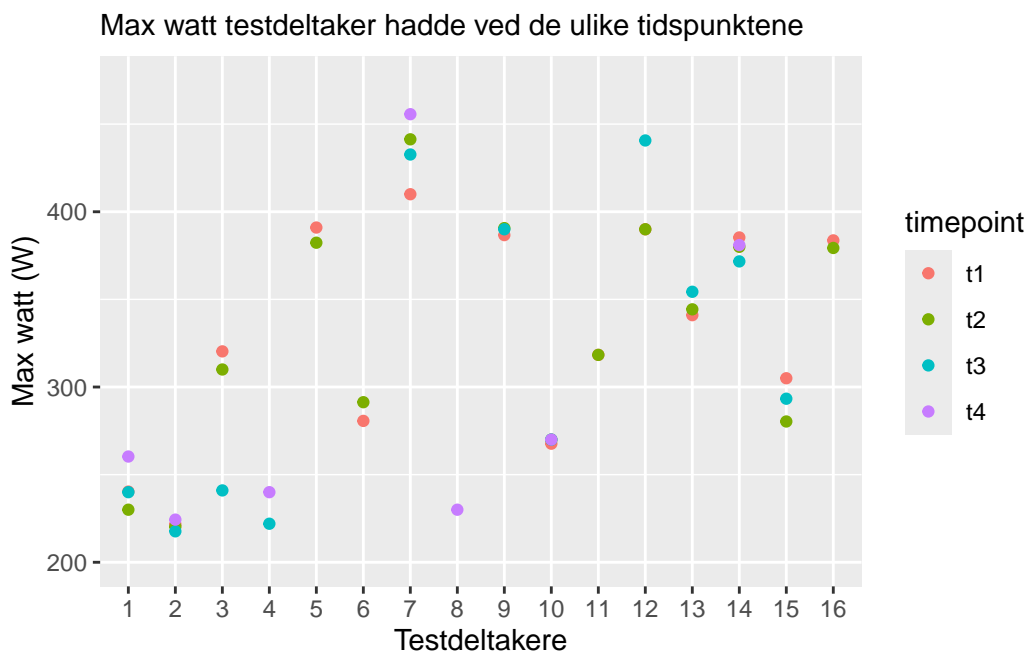
1.4 Intern validering

Både tekniske og biologiske faktorer kan potensielt påvirke resultatene og true den interne validiteten av testen (Halperin, Pyne, and Martin 2015). Ved å beskrive og etablere testprosedyre samt standardisere testen, ønsker vi å skape rammer som sikrer at vi kontrollerer for mulige konfunderende faktorer som kan påvirke testprestasjon. Dette vil bidra til å redusere variasjoner og sikre at resultatene reflekterer faktiske prestasjoner, snarere enn påvirkninger fra ytre faktorer.

Selv om vi har jobbet for å sikre god intern validitet, er det faktorer som vil kunne påvirke resultatene grunnet individuelle forskjeller i måten å opptre som testleder, bruk av musikk, samt humør og mental status til testdeltager (Halperin, Pyne, and Martin 2015).

1.5 Tabeller og plots med datasett fra alle gruppene

1.5.1 Plot



1.5.2 Tabell

Table 1.1: Verdier er gj.snitt og standardavvik for de ulike testtidspunktene

timepoint	w.max
t1	332 (62)
t2	331 (66.6)
t3	316 (84.7)
t4	294 (88.8)

1.5.3 Beregning av reliabilitet

Table 1.2: Sammenlikner t1 og t2 og ser på gj.snittsverdi, st.avvik, typ. error og coeff. of variation av variabelen w.max

m	s	te	cv
331.07	12.59	8.90	2.69

Ifølge Hopkins (2000) er “typical error” den gjennomsnittlige variasjonen ved en test. Typical error gir en indikasjon på den forventede variasjonen ved en retest. I dette tilfellet hvor vi ser på watt.max, kan en forvente at watt.max ved neste test vil variere med 2.69 % (Hopkins 2000).

2 Assignment 2: Regression models, predicting from data

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 analysemtoden 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 (Pisică et al. 2022).

En lineær regresjonsmodell er en modell der en kan estimere verdien av en avhengig variabel basert på verdien av andre kjente uavhengige variabler (Pisică et al. 2022). 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:

$$y_i = b_0 + b_1x_i + e_i$$

der y_i er den avhengige variabelen som kan estimeres ved å benytte de uavhengige variablene b_1x_i og b_0 . b_0 er skjæringspunktet til grafen og b_1 er stigningstallet til grafen.

2.2 Part 1 - Lactate thresholds

2.2.1 Metode

Dataene ble sortert i mer passende format (tidydata) for å gjøre det lettere med modellene våre senere. Deretter ble det gjort ulike regresjonsmodeller for å fremstille dataene våre. Nye skjæringspunkter ble tegnet for å illustrere treningsintensitet ved ulike laktatverdier.

2.2.2 Resultat

```

library(readxl)
library(tidyr)
library(exscidata)
library(tidyverse)
library(ggplot2)
library(dplyr)
library(magrittr)
library(gt)
library(ggtext)

# Kalkuler treningsintensitet ved 2 og 4 mmol/L

w <- cyclingstudy |>
  select(subject, group, timepoint, lac.225:lac.375) |>
  filter(timepoint == "pre", subject == 10) |>
  pivot_longer(names_to = "watt",
               values_to = "lac",
               names_prefix = "lac.",
               names_transform = list(watt = as.numeric),
               cols = lac.225:lac.375)

mod <- lm(lac ~ watt, data = w)

w |>
  ggplot(aes(watt, lac, group = subject)) +
  labs(x = "Watt",
       y = "Laktat") +
  geom_point(size = 3, shape = 21, fill = "gold") +
  geom_line(lty = 2) +
  geom_smooth(method = lm, se = FALSE) +
  geom_smooth(method = lm, se = FALSE, formula = y ~ poly(x, 2), color = "lightgreen") +
  geom_smooth(method = lm, se = FALSE, formula = y ~ poly(x, 3), color = "lightpink") +
  coord_cartesian(xlim = c(220, 380)) +
  geom_hline(yintercept = 2, color = "red") +
  geom_hline(yintercept = 4, color = "purple") +
  # Legger inn en vertikal linje, med y = 2mmol, skjæringspunkt 308W, tatt på øyemål.
  geom_vline(xintercept = 309, linewidth = 1, alpha = 0.8) +
  # Legger inn en vertikal linje med y = 4mmol, skjæringspunkt 342W, tatt på øyemål.
  geom_vline(xintercept = 342, linewidth = 1, alpha = 0.8) +
  theme_minimal()

```

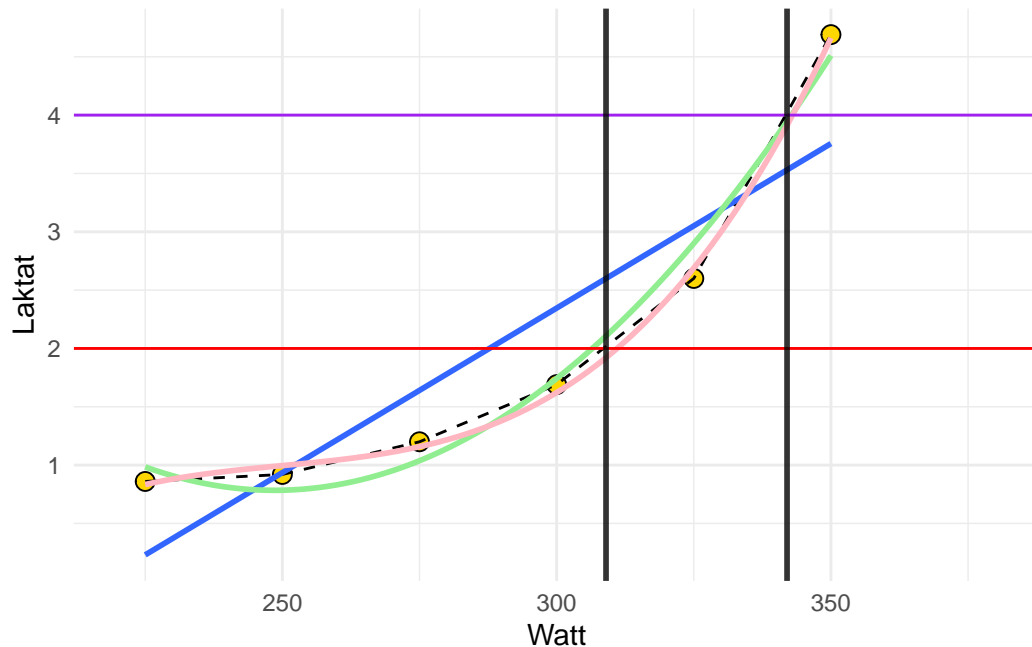


Figure 2.1: Figur 1: Gule punkter = laktat og watt, blå linje = lineær regresjon, grønn linje = andregradsligning, rosa = tredjegradsligning.

```
x2lac <- (2 - coef(mod)[1]) / coef(mod)[2]
x4lac <- (4 - coef(mod)[1]) / coef(mod)[2]
```

2.2.3 Diskusjon

Vi har valgt å se på subject 10 fra datasettet Cyclingstudy. Vi gjør om datasettet til tidy-data. 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 $1.69 \text{ mmol} \times \text{L}^{-1}$.

2.3 Part 2 - Predicting sizes of DNA fragments

2.3.1 Metode

For å kunne predikere kalibreringskurven til qPCR, er det en rekke prosesser på molekylærlabben som er gjort på forhånd før det kunne kjøres i R-studio.

For å kunne kjøre et PCR på en agrose gel 2%, ble det først tatt helbod fra en forsøksperson for å kunne ha et ekstrahert DNA. Helblodet vært igjen ulike prosesser hvor det er tilsatt ulike løsninger og tilsatt ulike primere. Vi står da igjen med et PCR-produkt. Det har videre blitt kjørt en elektroforese for å kunne separere DNA fragmentene fra PCR reaksjonen. Når elektroforesen var ferdig, ble det tatt et bilde av agrose gel 2%.

Videre har bildet fra elektroforesen blitt analysert ved hjelp av ImageJFiji og videre dataanalyse gjort i R og R studio. PCR-reaksjoner blir bestemt av primerdesign og dens spesifisitet.

2.3.2 Resultat

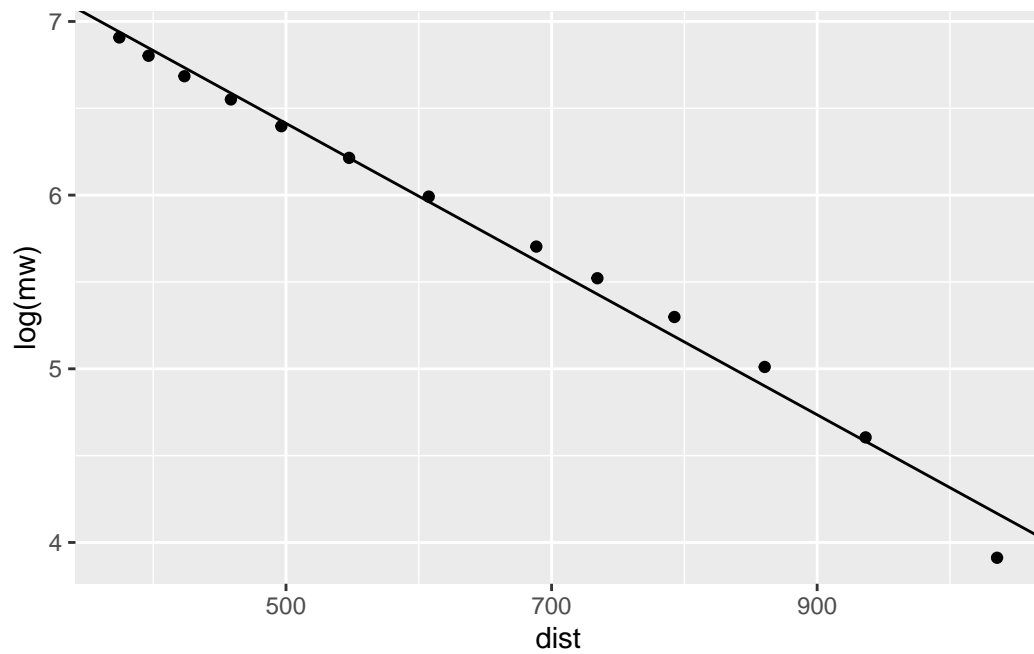
```
ladder <- data.frame(dist = c(374.5, 396.5, 423.5, 458.5, 496.5, 547.5, 607.5, 688.5, 734.5, 792.5),
                    mw = c(1000, 900, 800, 700, 600, 500, 400, 300, 250, 200, 150, 100, 50))

# Create a new data frame of unknowns
unknown <- data.frame(dist = c(1208.5, 600.5, 18.5, 383.5, 408.5, 436.5, 470.5, 508.5, 559.5))

# Fit the model
cal <- lm(log(mw) ~ dist, data = ladder)

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

ladder %>%
  ggplot(aes(dist, log(mw))) +
    geom_point() +
    geom_abline(intercept = coef(cal)[1], slope = coef(cal)[2])
```



862.5	133.29586
935.5	98.13646
993.5	76.94317

2.3.3 Diskusjon

2.4 Part 3 - Interpreting a regression table

2.4.1 Metode

2.4.2 Resultat

```
library(exscidata)
library(tidyverse)

dat <- hypertrophy |>
  select(GROUP, TRAINING_AGE, AGE, CLUSTER, VL_T2, TESTOSTERONE_T1, TESTOSTERONE_T2)

m <- lm(TESTOSTERONE_T1 ~ TRAINING_AGE, dat)

dat |>
  ggplot(aes(TRAINING_AGE, TESTOSTERONE_T1)) +
    labs(x = "Treningsalder (år)",
         y = expression(Testosteronverdier ~ ng %.% dl^{-1})) +
    # use expression for superscript

    geom_point(size = 3, shape = 21, fill = "orange") +

    geom_abline(intercept = coef(m)[1], slope = coef(m)[2], color = "steelblue", size = 1) +

    geom_hline(yintercept = coef(m)[1] + coef(m)[2] * 10, color = "darkgreen") +

    geom_vline(xintercept = 10, color = "darkred") +

    scale_y_continuous(breaks = c(200, 354, 400, 600, 800, 1000),
                       labels = c(200, "testo10", 400, 600, 800, 1000)) +

    theme_bw()
```

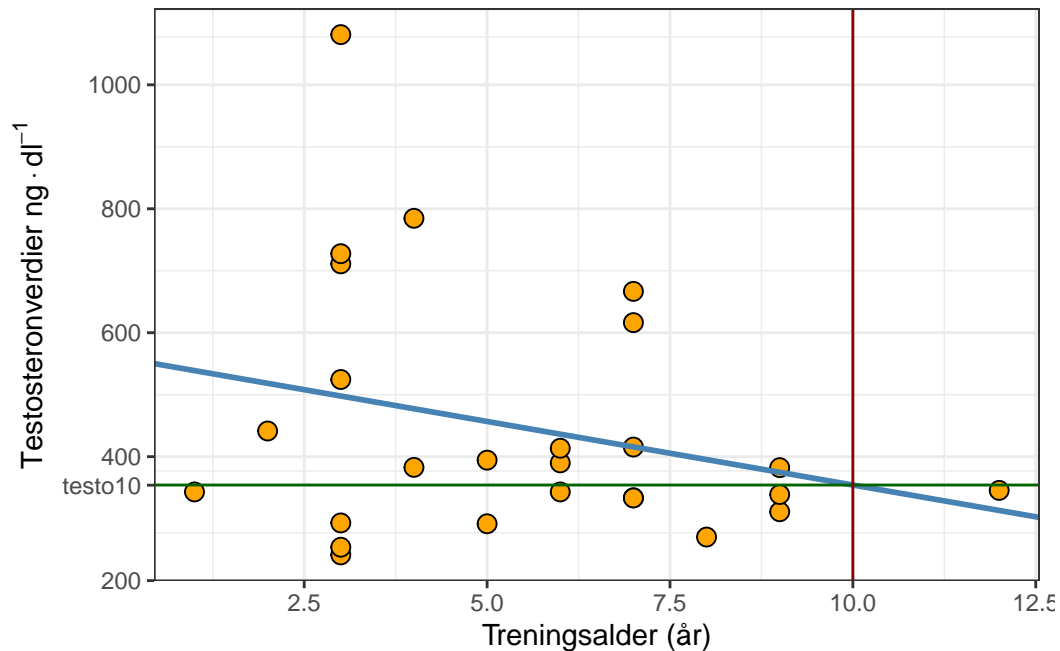


Figure 2.2: Figur 3: Sammenheng mellom treningsalder og testosteronverdier i blodet

```
# Den lineære modellen forteller i dette tilfellet at for hvert år man trener, så vil nivået av testosteron i blodet synke med 20.51 ng · dl⁻¹

testo10 <- coef(m)[1] + coef(m)[2] * 10
testo10rounded <- round(testo10, 2)

# testosteronnivå etter 10 år med trening estimeres til 354.26 ng·dl⁻¹
```

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 0,1779, 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 6.250. Den høye t-verdien på 6.250, og en p-verdi på 0,1779. 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.

2.5 Referanser

3 Assignment 3: Drawing inference from statistical models, and statistical power

This assignment is set up as a statistical laboratory, we will perform simulations and your assignment is to interpret and explain the results. Create a report based on the code used in the lab and make sure you answer the specified questions (1-8). You can be as creative as you want and explore the results further.

4 Assignment 4: Study designs

4.1 Overview

Choose an area of interest (e.g. protein supplementation for muscle hypertrophy or the effect of block periodization on VO2max). Find at least five *original research studies*¹ in your selected area and describe strength and weakness of these studies. The report should focus on the design of the studies and selection of statistical tests to answer study aims. Conclude your report with a recommendation, how should future studies in your area be designed to best answer similar questions?

¹Avoid using review articles or meta-analyses

5 Assignment 5: Analyzing repeated measures experiments

5.1 Assignment overview

In this assignment you will analyse and report on trial investigating the effect of resistance training volume on lean mass and muscle strength. The data are part of the `exscidata` package and can be accessed as `data("strengthvolume")` and `data("dxadata")`. Read the [instructions carefully!](#)

Below you will find a basic outline of the report and example code that we worked on in class.

5.2 Introduction

5.3 Methods

5.3.1 Participants and study overview

5.3.2 Muscle strength and hypertrophy

5.3.3 Data analysis and statistics

5.4 Results

The average difference in lean mass changes between sets were 122.8, 95% CI: [8.6, 237], $p = 0.036$.

```
## Time points in strength data set

strengthvolume %>%
  distinct(exercise)
```

```
# A tibble: 6 x 1
  exercise
  <chr>
1 legpress
2 legext
3 isok.60
4 isok.120
5 isok.240
6 isom
```

```
## Exploratory plot of strength data
```

```
str <- strengthvolume %>%
  filter(include == "incl") %>%
  mutate(time = factor(time, levels = c("pre", "session1",
                                         "week2", "week5",
                                         "week9", "post"))) %>%
  print()
```

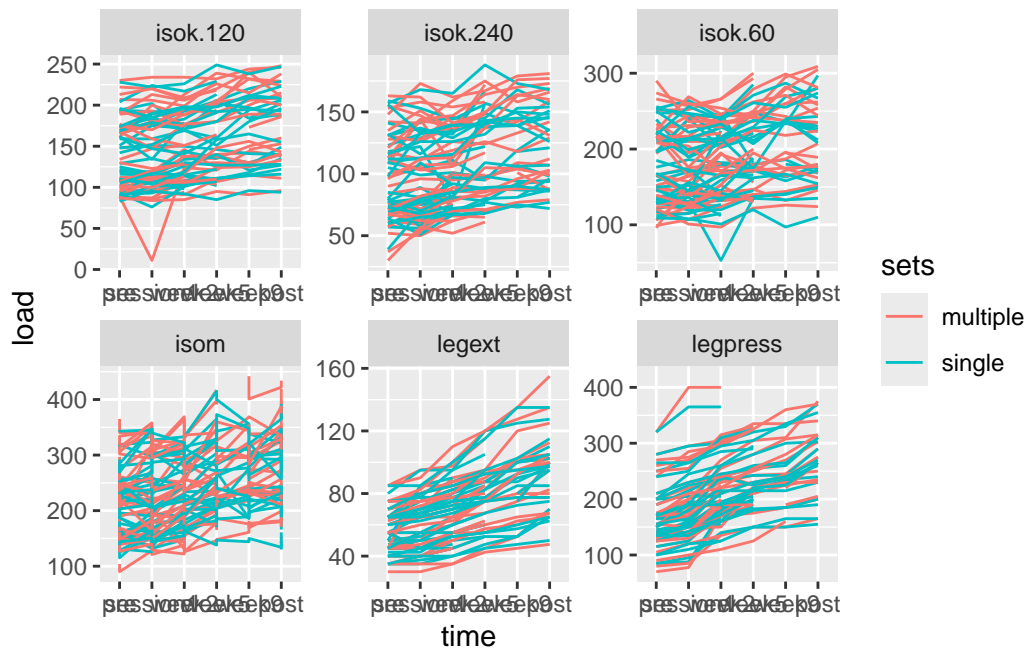
```
# A tibble: 2,856 x 8
```

	participant	sex	include	time	sets	leg	exercise	load
	<chr>	<chr>	<chr>	<fct>	<chr>	<chr>	<chr>	<dbl>
1	FP13	male	incl	pre	single	R	legpress	115
2	FP13	male	incl	pre	multiple	L	legpress	115
3	FP13	male	incl	pre	single	R	legext	55
4	FP13	male	incl	pre	multiple	L	legext	55
5	FP13	male	incl	session1	single	R	legpress	125
6	FP13	male	incl	session1	multiple	L	legpress	125
7	FP13	male	incl	session1	single	R	legext	55
8	FP13	male	incl	session1	multiple	L	legext	55
9	FP13	male	incl	week2	single	R	legpress	185
10	FP13	male	incl	week2	multiple	L	legpress	175

```
# i 2,846 more rows
```

```
str %>%
  ggplot(aes(time,
              load,
              group = paste(participant, sets),
              color = sets)) +
  geom_line() +
  facet_wrap(~ exercise, scales = "free")
```

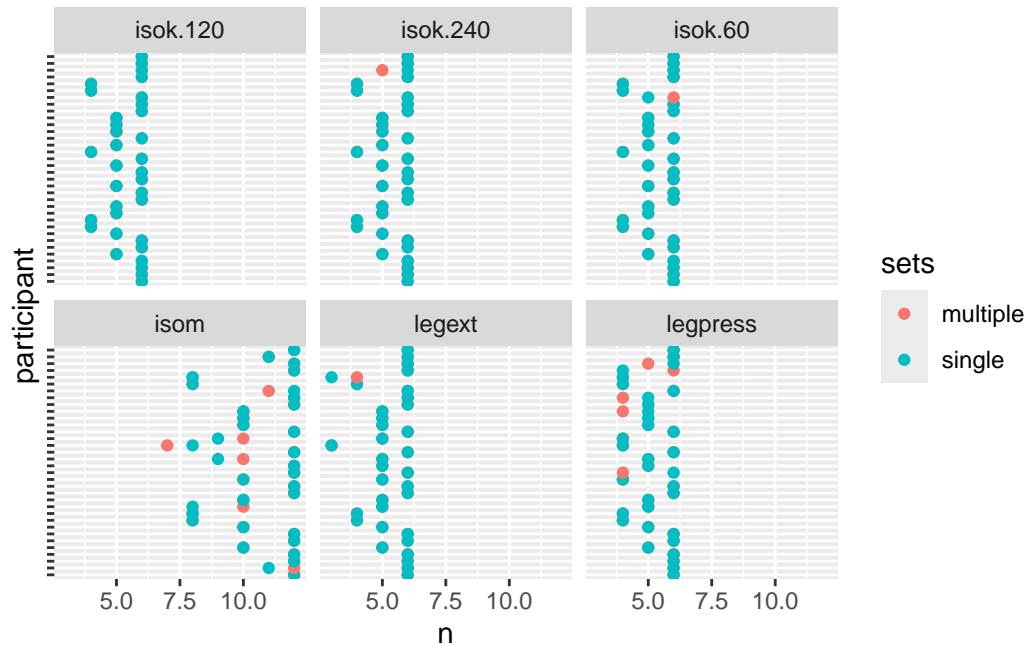
Warning: Removed 5 rows containing missing values or values outside the scale range (`geom_line()`).



How many measurements per participant

```
str %>%
  filter(!is.na(load)) %>%
  group_by(participant, exercise, sets) %>%
  summarise(n = n() ) %>%
  ggplot(aes(n, participant, color = sets)) +
  geom_point() +
  facet_wrap(~ exercise) +
  theme(axis.text.y = element_blank())
```

`summarise()` has grouped output by 'participant', 'exercise'. You can override using the `.groups` argument.



```
## Use pre and post data
# Combine pre data prior to data analysis
# per exercise, leg, participant, and sets

str %>%
  mutate(time = if_else(time %in% c("pre", "session1"), "pre", time)) %>%

  filter(time %in% c("pre", "post")) %>%

  summarise(load = max(load, na.rm = TRUE),
            .by = c(participant,
                    sex,
                    time,
                    sets,
                    exercise,
                    leg)) %>%

  print()
```

Warning: There were 7 warnings in `summarise()`.
 The first warning was:
 i In argument: `load = max(load, na.rm = TRUE)`.

```
i In group 62: `participant = "FP6"`, `sex = "female"`, `time = "post"`, `sets
  = "multiple"`, `exercise = "legpress"`, `leg = "L"`.
Caused by warning in `max()``:
! no non-missing arguments to max; returning -Inf
i Run `dplyr::last_dplyr_warnings()` to see the 6 remaining warnings.
```

```
# A tibble: 816 x 7
  participant sex    time sets    exercise leg    load
  <chr>      <chr> <chr> <chr>    <chr>    <chr> <dbl>
1 FP13      male   pre   single  legpress R      125
2 FP13      male   pre   multiple legpress L      125
3 FP13      male   pre   single  legext   R       55
4 FP13      male   pre   multiple legext   L       55
5 FP13      male   post  single  legpress R     230
6 FP13      male   post  multiple legpress L     235
7 FP13      male   post  single  legext   R     97.5
8 FP13      male   post  multiple legext   L     100
9 FP16      female pre   single  legpress R      95
10 FP16     female pre   multiple legpress L      85
# i 806 more rows
```

5.5 Discussion

5.6 Conclusion

6 Philosophy of science

See instructions on canvas.

7 Molecular Laboratory report

Select one laboratory assignment and write a detailed report.

References

- Halperin, Israel, David B Pyne, and David T Martin. 2015. “Threats to Internal Validity in Exercise Science: A Review of Overlooked Confounding Variables.” *Int. J. Sports Physiol. Perform.* 10 (7): 823–29.
- Hopkins, W G. 2000. “Measures of Reliability in Sports Medicine and Science.” *Sports Med.* 30 (1): 1–15.
- Pisică, Dana, Ruben Dammers, Eric Boersma, and Victor Volovici. 2022. “Tenets of Good Practice in Regression Analysis. A Brief Tutorial.” *World Neurosurg.* 161 (May): 230–239.e6.