IDR3002 - 2019 - MAPPEEKSAMEN

Emne	Emnekode	Semester
Kvantitativ metode og statistikk	IDR3002	19H

Læringsutbytte

Kunnskap:

- Kandidaten kan vurdere ulike forskningsdesign som er aktuelle innen fagområdet
- Kandidaten kan vurdere den metodiske kvaliteten i ulike studiedesign
- *Kandidaten kjenner til etiske prinsipper innen medisinsk forskning og regelverk for forskning på mennesker*
- Kandidaten kan vurdere kvantitative metoder som benyttes for å innhente data om fysisk prestasjonsevne og helse, med særlig fokus på fysiologiske, cellebiologiske og molekylærbiologiske egenskaper
- *Kandidaten har kjennskap til sentrale aspekter ved helse, miljø og sikkerhet i forbindelse med laboratoriearbeid*
- Kandidaten kan vurdere de statistiske metodene som er mest aktuelle innen fagområdet.

Ferdigheter:

- Kandidaten kan lete fram vitenskapelige kunnskapskilder og bruke disse på hensiktsmessig måte innen oppgaveskriving og faglige diskusjoner
- Kandidaten kan planlegge og gjennomføre innsamling av kvantitative vitenskapelige data og presentere skriftlig analyse av dataene på en selvstendig måte
- *Kandidaten kan utføre arbeidsfysiologisk testing*
- *Kandidaten kan utføre cellebiologiske og molekylærbiologiske analyser under veiledning*
- Kandidaten kan behandle innsamlede data på en ryddig og etterprøvbar måte
- Kandidaten kan gjennomføre de statistiske analysene som er mest aktuelle innen fagområdet i et dataprogram

Generell kompetanse:

- Kandidaten kan vurdere utvalgte vitenskapelige metoder, anvende disse og stå for praktisk datainnsamling i nye prosjekt
- Kandidaten kan planlegge og gjennomføre de statistiske analyser som er aktuelle for fagområdet.
- Kandidaten kan planlegge og gjennomføre dataanalyse på en slik måte at det blir reproduserbart og fremstilling av forskningsresultater skal være etterprøvbare.
- Kandidaten kan uttrykke seg skriftlig i henhold til fagområdets egenart og utforme vitenskapelige rapporter under veiledning.

Eksamensoppgaver med prinsipper for vurdering

The exam in the course IDR3002 is a portfolio examination. With the deadline set to 2019-12-06 you are expected to hand in the following as your portfolio:

- A short report covering the "Reliability study"
- Statistical assigment 1
- Statistical assigment 2
- Statistical assigment 3
- Statistical assigment 4
- Statistical assigment 5
- Laboratory report

The portfolio exam is graded pass/fail.

A short report covering the "Reliability study"

Analyze the data from the reliability study and create a revised report based on your preliminary findings (reported in a presentation on 2019-09-06). Keep it short and concise (maximum 1 page text). Your revised analyses is preferrably done in R!

Statistical assigment 1

In this assignment you are asked to reproduce at least 3 of 4 figures. We will be using the cycling data set and ggplot2 together with dplyr and tidyr. The graphs to reproduce are shown below together with instructions for each graph.

Figure 1: Correlation between percentage change in counter movement jump and peak power

Anaerobic performance might be affected by endurance training, to see if changes in performance in a non-specific exercise (counter-movement jump) is related to changes in cycling-specific peak-power we are going to plot the two.

Relevant variables from the `cyclingStudy` data set are

Variable	Description
`subject`	ID of participants
`timepoint`	Time-point for testing pre-training (`pre`), and after each training period (`meso1-meso3`)
`cmj.max`	Counter movement jump performance
`peak.power`	Peak power from a Wingate test

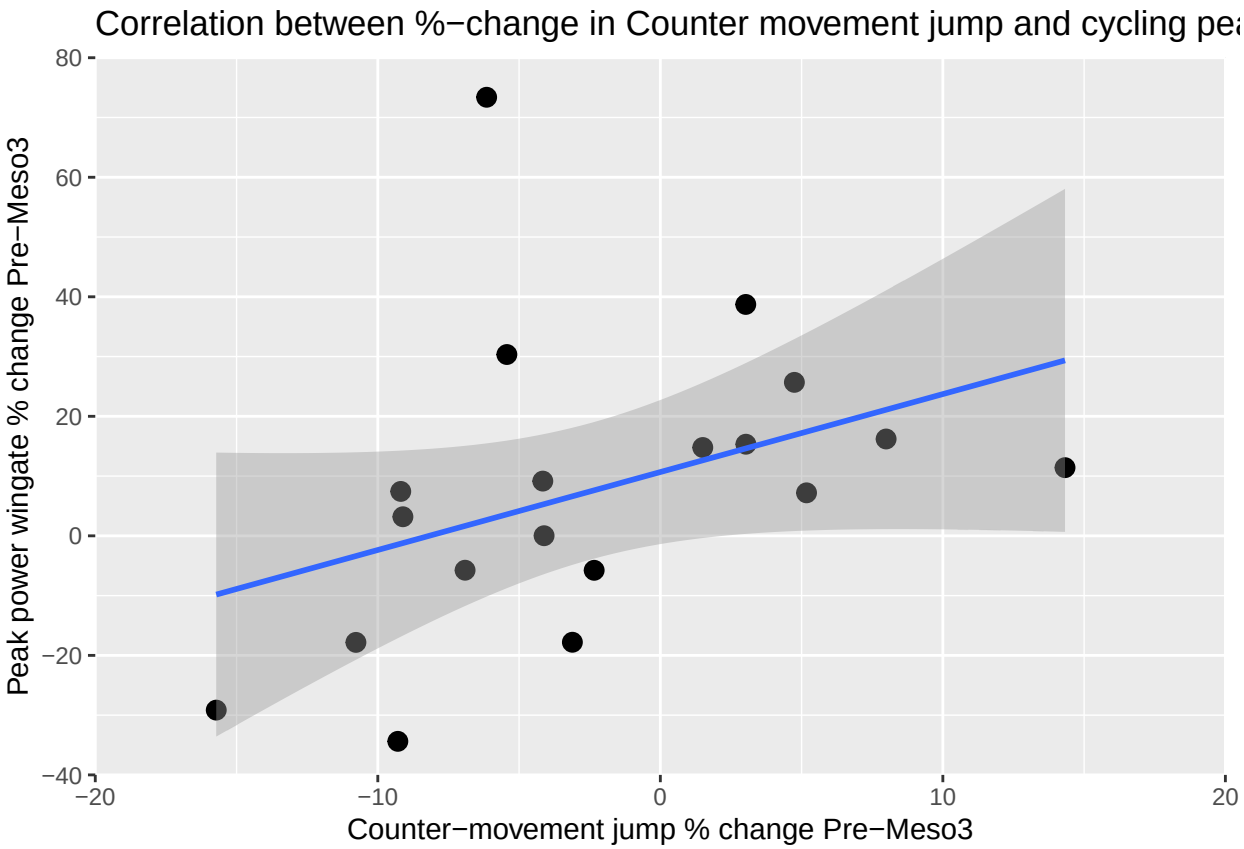


Figure 2: Group-wise change in VO_{2max} ml^{-1} kg^{-1} min^{-1}

In the figure below, we display percentage increase in VO_{2max} (ml^{-1} kg^{-1} min^{-1}) per group. Relevant variables from the `cyclingStudy` data set are:

Variable	Description
`subject`	ID of participants
`timepoint`	Time-point for testing pre-training (`pre`), and after each training period (`meso1-meso3`)
`group`	ID of the groups, `INCR`, `DECR` and `MIX`
`weight.T1`	Body mass
`VO2.max`	VO_{2max} in ml^{-1} min^{-1}

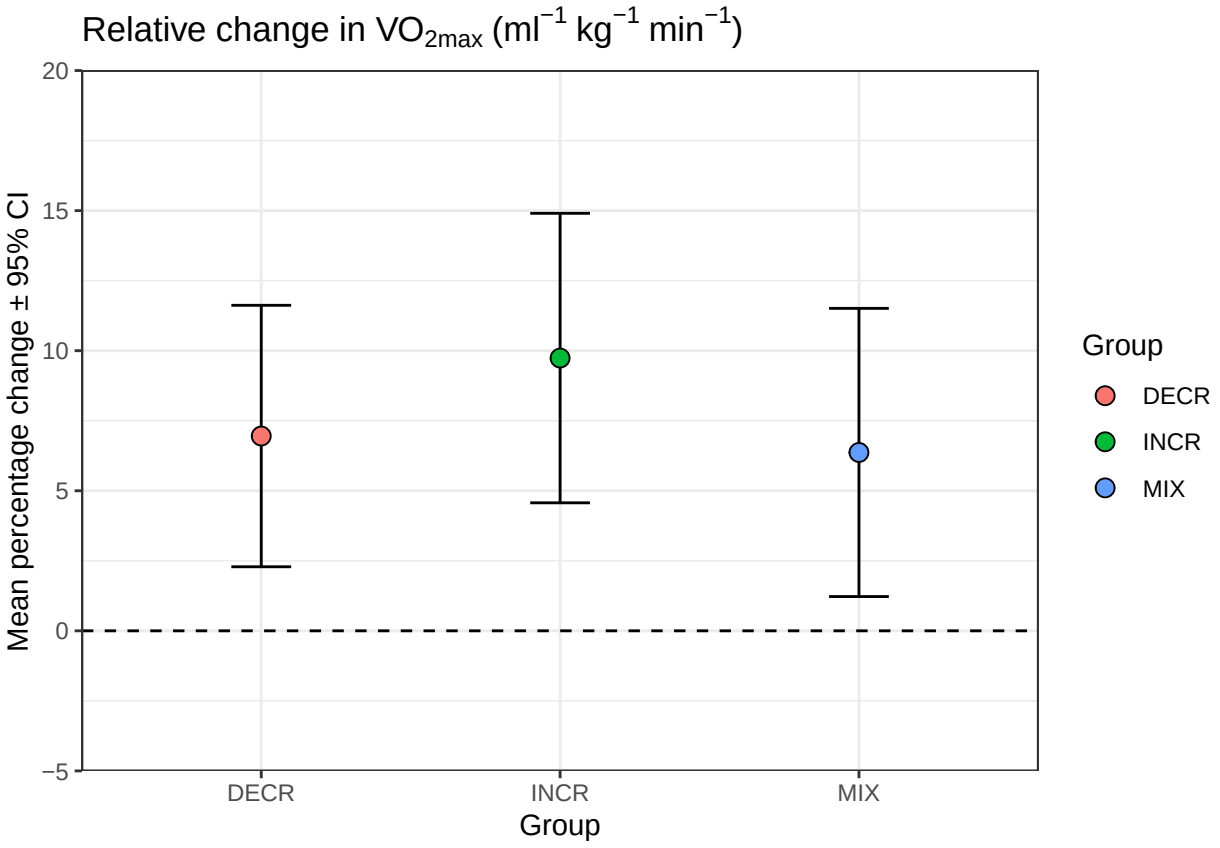


Figure 3: Faceting the correlation between VO_{2max} and sub maximal VO_2.

To display data in different panels can be useful when doing exploration graphics. Here we will plot maximal VO_2 against sub maximal VO_2 per test and watt.

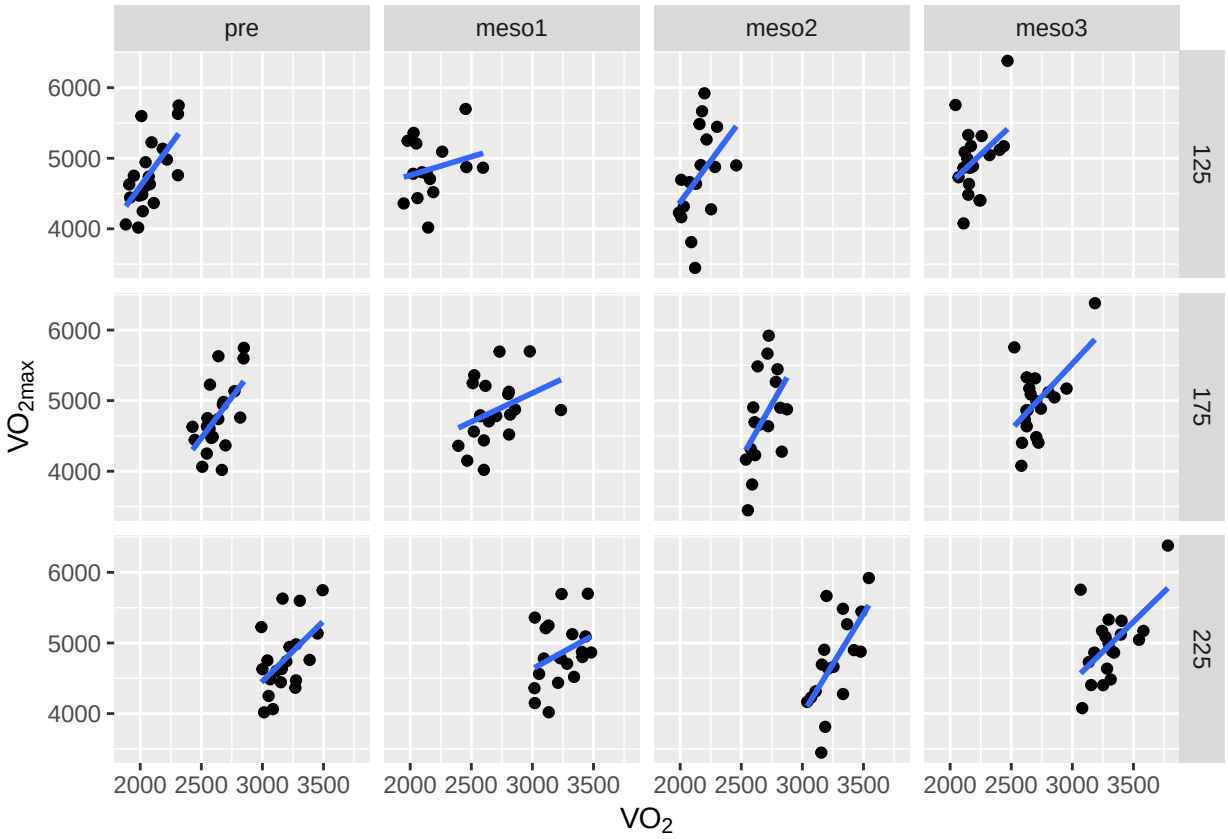
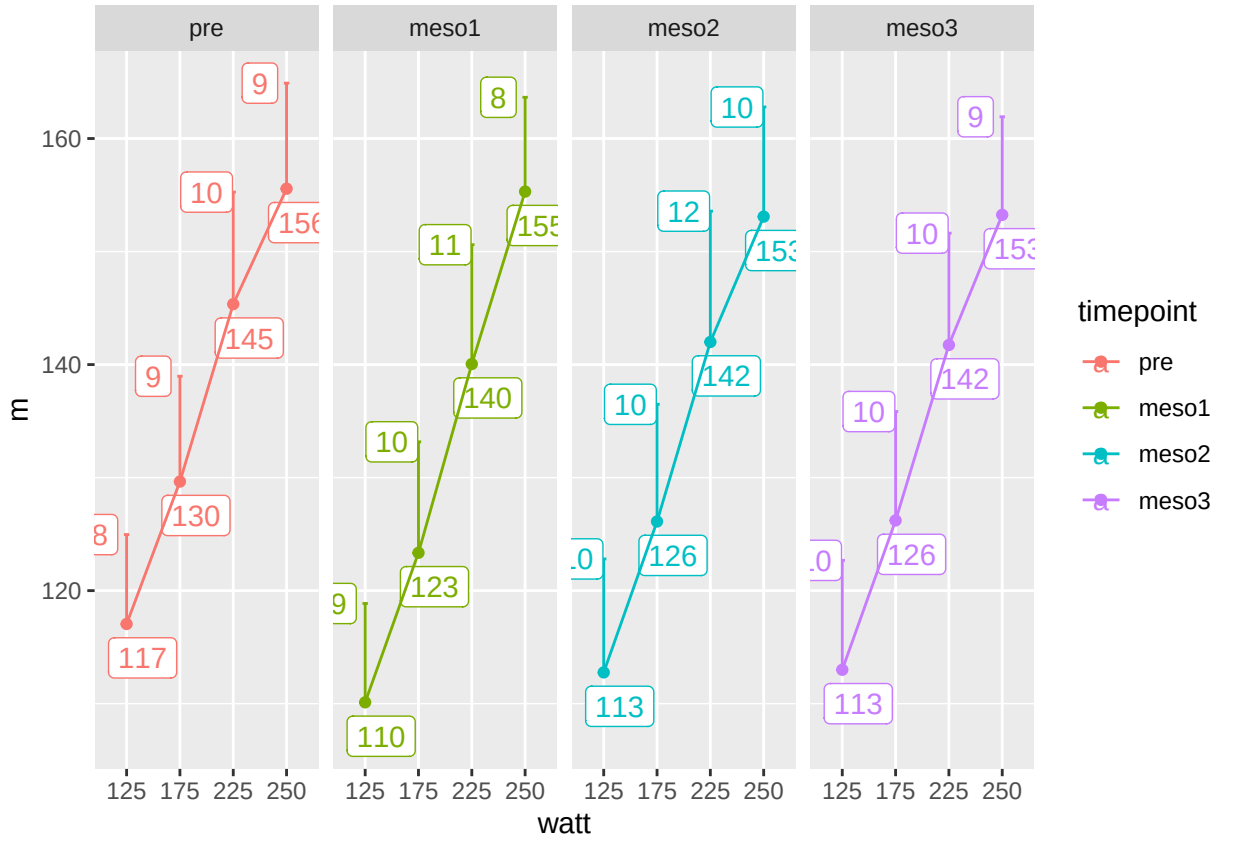


Figure 4: Mean and standard deviation of multiple variables

We will now use sub maximal heart rate values (`HF.125` to `HF.250`) and plot these against watt. We are interested in the mean and standard deviation per time point.



Statistical assigment 2

This week you are expected to draw conclusions about the "real world" based on some data we have collected.

The html document should not display code chunks, messages or warnings. This means that you should report your results in ordinary text and include results there with inline code. Read more about this here.

You are expected to answer all three problems below.

Problems

1. Using the cyclingStudy data set, what can you say about the effect of training on VO_{2max} per kg body mass. Here you should test against the H_0 that says that VO_{2max} is no different between `pre`and `meso1`. Interpret the results of your test. What test do you use? Where do you set the level of required evidence to reject the null, and why?

2. Create a new variable where you calculate the difference in `sj.max` in percentage between `pre` and `meso1`. Do a test against the null-hypothesis that this difference is 0. What test do you use. What is your threshold to reject the null?

3. Calculate the percentage change in VO_{2max} in group `INCR` and `DECR`between `pre` and `meso3`. Compare the two groups, what statement can you make about the population on the basis of your test? What test do you choose, why?

Intepretations of the results should contain point estimates, confidence intervals and p-values.

Statistical assigment 3

This week you are expected to perform statistical analysis using regression models (hide the code in the report).

Problem 1

Using the cycling data set. Fit the following model:

$$VO_{2max} = Height$$

Use data from time-point `meso2`. Check your assumptions and interpret the model.

Problem 2

Phosphorylation of the p70 S6 kinase induces increased translational efficiency. In this study (Mitchel et al. 2013), the researcher asked if baseline exercise-induced phosphorylation of p70S6K could explain changes in Cross Sectional Area (CSA) of muscle fibers after a training period.

Methods: Young male participants (n=23) completed a 16 wk training period with 4 weekly strength training sessions. Muscle biopsies were collected from from vastus lateralis before and after the training period. Two biopsies were collected before the training period, before and after an acute exercise session in order to study activation of molecular pathways involved in protein synthesis. Muscle fiber CSA were analyzed from muscle tissue cross sections.

Data analysis questions: 1. What is the relationship between phosphorylation status and increase in CSA? 2. How much of the variation in CSA can be explained by p70 S6K? 3. Examine the model fit and check assumptions, identify potential outliers and compare your conclusions to the conslusions reported in the paper.

Reference: Mitchell, C. J., et al. (2013). "Muscular and Systemic Correlates of Resistance Training-Induced Muscle Hypertrophy." PLoS One 8(10): e78636.

Problem 3

Using the cycling data set. Select a performance variable from the list

- `VO2.max` (VO_{2max})
- `peak.power` (peak power in wingate test)
- `tte` (time to exhaustion in VO_{2max} test)
- `sj.max`

Find 2-4 variables that might explain the variation in one of these performance variables. Fit a full model (using all of your selected variables). Remove variables that do not add information to the model. The goal is to explain as much as possible of the performance.

You might use p-values, R^2 or the `AIC()` function to select variables.

Check your assumptions and interpret the final model.

For this assignment you are expected to use the course literature to motivate your work-flow.

Diez, D. M., C. D. Barr, and M. Çetinkaya-Rundel. 2015. OpenIntro Statistics. OpenIntro, Incorporated

Navarro, Daniel. 2019. Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners (Version 0.6.1)

Statistical assigment 4

Pre- to post-exercise trial may be analyzed in many different ways. This week we have looked at t-test of change-scores, ANCOVA with baseline correction and change scores as outcome and mixed effects models on raw scores.

The report shoul contain:

- Introduction: Using at least three references, describe the field of training volume and strength gains. State the objective of your study.
- Methods: Use the information you have about the study and write a short methods section about how the data (or how you think the data was collected). Write a **Statistics** section to describe how you analyzed the data.
- Results: The results should contain estimates of group differences, a figure to show trends over time and numbers used to draw inference (p-values or confidence intervals). Use mixed models or ANCOVA to estimate differences.
- Discussion: A short section where you interpret your results in relation to the studies you have presented in the introduction.

Use the strengtTests.csv data set. A more detailed description of the study can be found here:

Daniel Hammarström, Sjur Øfsteng, Lise Koll, Marita Hanestadhaugen, Ivana Hollan, William Apro, Jon Elling Whist, Eva Blomstrand, Bent R. Rønnestad, Stian Ellefsen bioRxiv 666347; doi: https://doi.org/10.1101/666347.

The data set was constructed from this study. You have analyzed only one leg per participant!

Statistical assigmnet 5 – Literature review: Study design and analysis

You are expected to write a short literature review (including 4-8 studies) with focus on **study designs** in a specific field. The title of your review should be "Research designs and analyses in studies examining X" where "X" is an exercise-physiology related problem. An example could be "exercise-induced protein synthesis" or "cardiovascular response to acute exercise". The first part of your review should focus on design. Your review should cover what aspects are common in all studies that you have examined and a discussion on why this might be the case. You should also highlight differences between studies. With regard to study design, you are also expected to discuss if the field is in need of new, better designed studies.

The second part of your review should examine statistical aspects of studies. Below is a list of possible questions that you can use to examine statistical aspects of selected studies.

- How many participants (or experimental units) are recruited to the study?
- Are there *a-priori* sample-size calculations?
- Are sample size calculations based on clinically important thresholds?
- What statistical tests are used to answer questions/hypotheses?
- Are statistical tests reported in full (estimates, p-values, test statistics)?
- Are assumptions about statistical tests discussed?
- Do conclusions match questions/hypotheses and statistical analyses?

References that might help you

Thiese (2014) provides an overview of study designs. Lakens (2015) discusses problems related to error rates in science. This paper provides you with a more specific discussion regarding statistical power and error rates extending the course literature (Navarro 2019; Diez, Barr, and Çetinkaya-Rundel 2015).

Laboratory report

Using results from laboratory exercises you are expected to write reports detailing analyses made in the course:

- Extraction of DNA from whole blood and Genotyping by PCR
- Extraction of RNA from muscle samples and Determination of relative mRNA abundance in muscle samples
- Extraction of Protein from muscle samples and Determination of Myosin heavy chain composition in skeletal muscle

The report should contain background information, methods, results and a short discussion/interpretation of the results. All experiments should be documented in your laboratory journal.

References

Diez, D. M., C. D. Barr, and M. Çetinkaya-Rundel. 2015. *OpenIntro Statistics*. OpenIntro, Incorporated. https://leanpub.com/openintro-statistics.

Lakens, D. 2015. "On the challenges of drawing conclusions from p-values just below 0.05." *PeerJ* 3: e1142.

Navarro, Daniel. 2019. *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners (Version 0.6.1)*. https://learningstatisticswithr.com/book/.

Thiese, M. S. 2014. "Observational and interventional study design types; an overview." *Biochem Med (Zagreb)* 24 (2): 199–210.