

# Bygge statistiske modeller med kategoriske prediktorvariabler (t-tester, ANOVA)

Christian Magelssen

2021-03-28



# Contents

1	Introduksjon	5
2	Datasett (3 sett vs. 1 sett)	7
3	Hvilken modell skal vi velge?	11
4	H0: Null-hypotese	13
5	H1: Alternativ hypotese	15
6	f-ratio og signifikanstesting	17
7	F-tester i Jamovi	19



# Chapter 1

## Introduksjon

I dette kapittelet skal vi lære å bygge statistiske modeller for å teste om **to eller flere grupper er forskjellige på en avhengig variabel som er kontinuerlig**.

En variabel sies å være **kontinuerlig** når vi kan bestemme hvor presist vi ønsker å måle det. For eksempel regnes tid som en kontinuerlig variabel fordi det (i prinsippet) ikke er noen grenser hvor fint vi kan måle tid; vi kan måle det i år, måneder, uker, dager, timer, minutter, sekunder, tideler, hundredeler eller tusendeler.

**Grupper** defineres innenfor psykologifaget som en samling mennesker som deler bestemte karakterstikker. Det kan være spillere på et fotballag, individer på et treningssenter, eller menn og kvinner. Dette er også eksempler på grupper der inndelingen faller seg naturlig. I eksperimentelle studier er disse.

La oss se på hvordan en slik analyse vil se ut.

Reference a figure by its code chunk label with the **fig:** prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from **knitr::kable()**, e.g., see Table 1.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2020) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

Table 1.1: Here is a nice table!

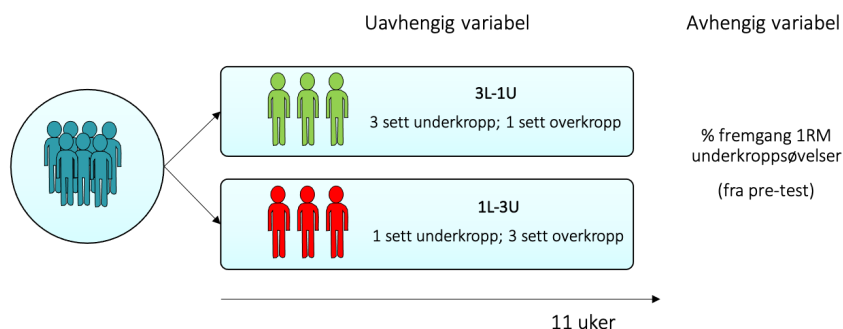
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

## Chapter 2

# Datasett (3 sett vs. 1 sett)

Et spørsmål mange treningsentusiaster lurer på er hvor mange serier som er best å gjennomføre for å få maksimal treningseffekt i styrketrening. Noen mener at ett sett er tilstrekkelig, mens andre mener at et hardere treningstimuli er nødvendig og at to eller flere sett derfor er bedre. En forsker som var tidlig ute med å undersøke dette er deres egen Bent Rønnestad.

Eksperimentet ble gjennomført som et **between-subject design** med to grupper: en gruppe trente 1 sett på underkroppen og 3 sett på overkroppen; En annen gruppe trente 3 sett på underkroppen og 1 sett på overkroppen. Disse gruppene kalte han henholdsvis **1L-3U** og **3L-1U** (L=lower; U=Upper). De to gruppene trente 3 ganger i uken i totalt 11 uker. Forskerne ville så se hva som ga best fremgang på 1RM. Den avhengige variabelen ble derfor %-fremgang på 1RM på underkroppsovelser. De fant at 3L-1U hadde større fremgang enn 1L-3U fra pre til post (41 vs 21 % endring). Denne forskjellen var signifikant ved en uavhengig t-test. Med andre ord kan det se ut til at det kan lønne seg å trene flere sett per styrketreningsøkt.



Vi har ikke tilgang til dette datasettet, men jeg har simulert dette datasettet i R basert på verdiene jeg fant i artikkelen. Datasettet blir tilnærmet likt,

Table 2.1: Simulert datasett

individ	gruppe	rm
1	tre.sett	40.46704
2	tre.sett	49.07223
3	tre.sett	47.94131
4	tre.sett	44.51389
5	tre.sett	52.28750
6	tre.sett	40.01750
7	tre.sett	49.48425
8	tre.sett	29.21048
9	tre.sett	40.59293
10	tre.sett	37.58676
11	tre.sett	35.42651
12	tre.sett	42.49354
13	ett.sett	17.70576
14	ett.sett	17.07181
15	ett.sett	18.26811
16	ett.sett	25.42594
17	ett.sett	32.70313
18	ett.sett	19.10226
19	ett.sett	22.23827
20	ett.sett	22.27148
21	ett.sett	26.17889
22	ett.sett	20.34857
23	ett.sett	23.52773
24	ett.sett	17.95966

men siden det er en simulering blir det aldri helt identisk. Datasettet ser du i tabellen under.

```
knitr::kable(  
  dat, booktabs = TRUE,  
  caption = 'Simulert datasett'  
)
```

Du kan få nøyaktig samme datasett ved å klippe ut og lime inn følgende kode i en skript-fil i R (husk å laste inn tidyverse-pakken, `library(tidyverse)`). Du kan også laste ned datasettet som en .csv fil fra canvas.



```
set.seed(2002) #viktig å ha med denne for å få nøyaktig samme datasett
tre.sett <- rnorm(n = 12, mean = 41, sd = 5) #12 individer
ett.sett <- rnorm(n = 12, mean = 21, sd = 5) #12 individer

#lager en tibble fra tidyverse-pakken. Må ha lastet inn tidyverse library(tidyverse) i scriptfilen
dat <- tibble(individ = seq(1:24),
              gruppe = rep(c("tre.sett ", "ett.sett"), c(length(tre.sett), length(ett.sett))),
              rm = c(tre.sett , ett.sett))
```

Før du går videre er det greit at du gjør deg kjent med datasettet som vi har generert. Studer datasettet og svar på følgende spørsmål:

1. Hvor mange kolonner er det i tabellen over?
2. Hvor mange deltakere var med i studien?
3. Hvilke to verdier kan variabelen gruppe? og

### 2.0.1 Regne gjennomsnitt for de to gruppene

Bra! Det er alltid viktig å bli kjent med sitt eget datasett, men nå som du har det kan vi gå videre. Vi er interessert i om det er forskjeller mellom de to gruppene (“tre.sett” vs. ett.sett) på % fremgang fra pre- til post-test. Så kanskje vi kan starte med å se om det er forskjeller i gjennomsnitt mellom to gruppene? Dette kan enkelt gjøres i R, Jamovi eller excel. Her er en kode for å gjøre dette i R:

```
#jeg lager et oobjekt som heter mean_rm
mean_rm <- dat %>%
  #Jeg grupperer etter gruppe, slik at jeg får et mean for hver gruppe istf. for å få mean for all
  #group_by er en funksjon for dette
  group_by(gruppe) %>%
  #deretter bruker jeg summarise funksjonen for å regne gjennomsnitt
  summarise(mean.fremgang.1RM = mean(rm))
```

Koden gir oss følgende tabell: \begin{table}

\caption{Gjennomsnittlige %-vis fremgang for de to gruppene}

gruppe	mean.fremgang.1RM
ett.sett	21.90013
tre.sett	42.42450

\end{table}

Hvilken gruppe hadde mest fremgang? mcq(c(“ett.sett”, answer = “tre.sett”))

## 2.0.2 Figur av datasettet

Vi kan også presentere dataen i en figur. En vanlig måte å gjøre dette på er å bruke et stolpediagram.

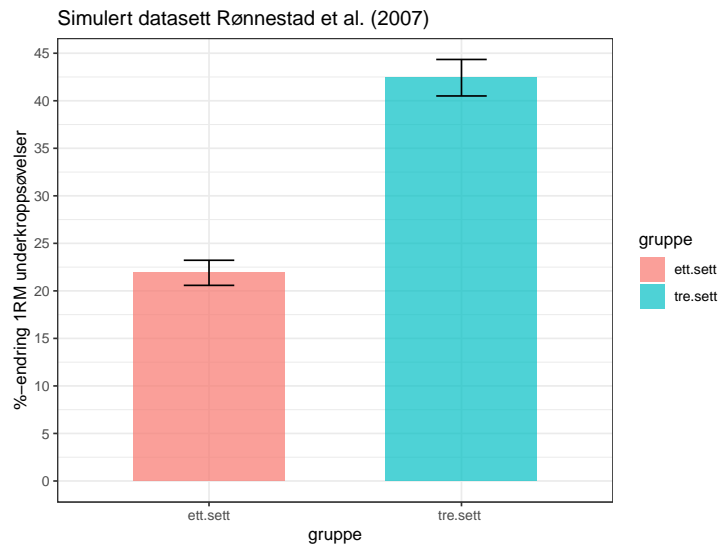


Figure 2.1: Here is a nice figure!

Et stolpediagram er pent å se på, men er egentlig designet for kategorisk data. For eksempel er det fint å bruke dette når vi skal presentere frekvensen antall som har kjørt bil til skolen og antall personer som har gått. Les Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm, og svar på følgende spørsmål:

```
knitr::include_app("https://cmagelssen.shinyapps.io/test/",  
  height = "600px")
```

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is
```

## Chapter 3

# Hvilken modell skal vi velge?

Her tenker jeg å si litt om at vi har ulike muligheter.



## Chapter 4

# H0: Null-hypotese

Her tenker jeg å si at Null-hypotesen er at det ingen forskjeller mellom gruppene. Og at alt vi trenger er å predikere et individs skår med gjennomsnittet for alle individene i hele datasettet.

Deretter regner jeg ut error for alle gruppene.



## Chapter 5

### H1: Alternativ hypotese

We have finished a nice book.





## Chapter 6

# f-ratio og signifikanstesting

We have finished a nice book.



## Chapter 7

# F-tester i Jamovi

We have finished a nice book.



# Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.21.