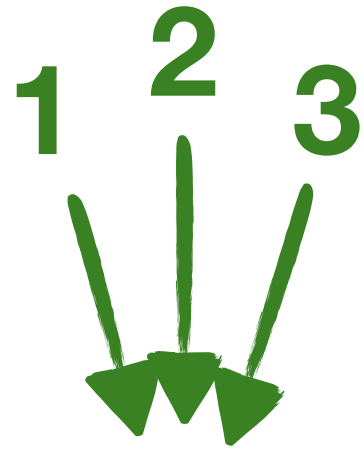




ISTT100y

Uke 1:

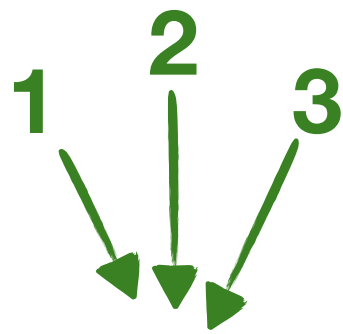
Beskrivende statistikk

Thea Bjørnland
Institutt for matematiske fag
NTNU

Uke 1: Beskrivende statistikk

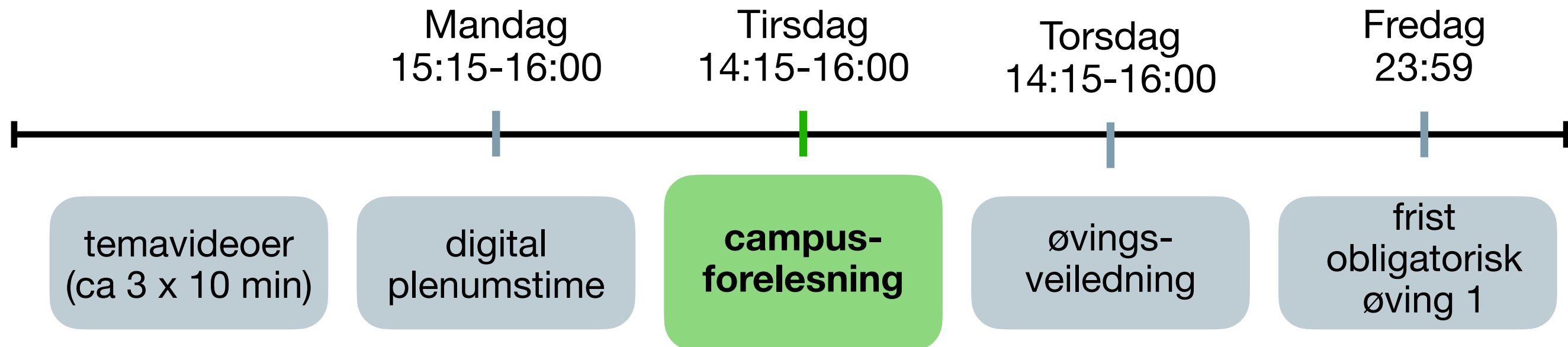
- Typer av data: kategoriske, diskrete, kontinuerlig
- Oppsummerende tall:
 - sentralmål (gjennomsnitt, median, ...)
 - mål på spredning (varians, standardavvik, ...)
- Figurer for å visualisere ulike typer data (histogram, boksplott, ...)
- Parvise målinger:
 - Kryssplott
 - Korrelasjon
 - Minste kvadraters rette linje

20. august 2024



ISTT100y uke 1

Beskrivende statistikk



I dag skal vi...

- Visualisere informasjon om dere og diskutere datatyper (mentimeter)
- Regne litt for hånd (oppsummerende tall)
- Diskutere populasjon og utvalg (mentimeter)
- Se på sammenhenger (korrelasjon) mellom ulike observasjoner

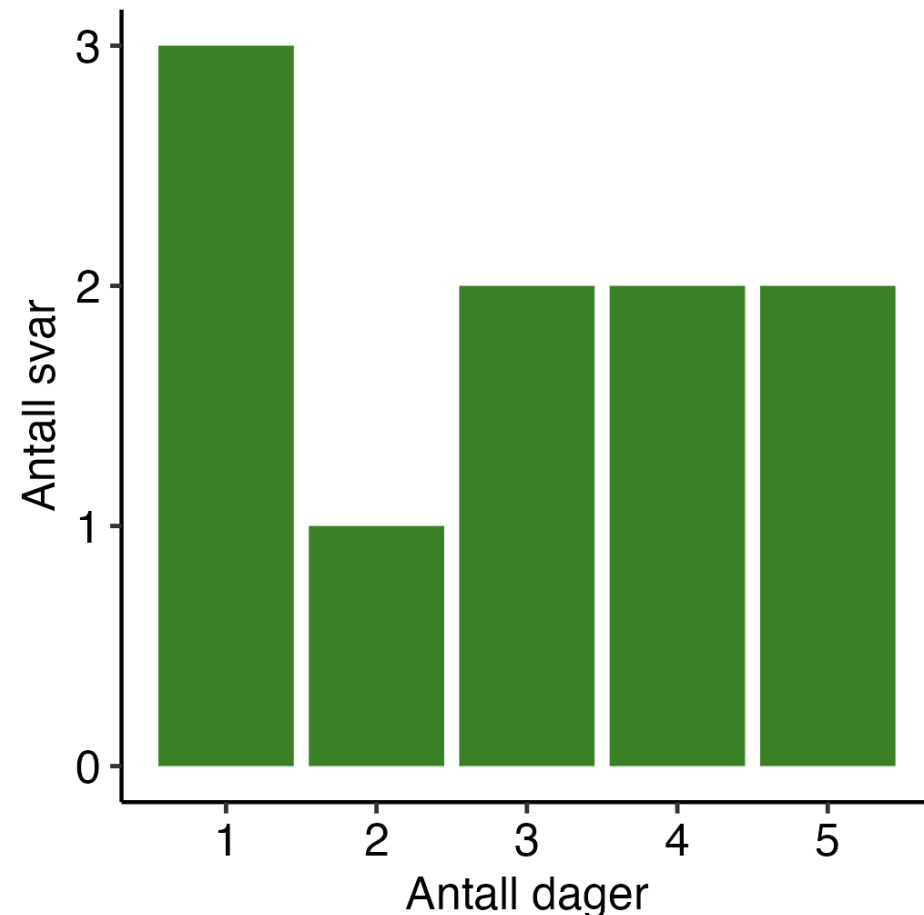
Tips! Alt jeg skal skrive på tavla ligger allerede ut på Blackboard og her: <https://wiki.math.ntnu.no/istx100y/2024h/istt>

Gå til www.menti.com

Oppsummerende tall

Vi regner for hånd

- Vi har spurt 10 kollegaer som sykler til jobb, hvor mange dager de syklet til jobb i forrige uke:
 - 1, 4, 4, 1, 3, 5, 2, 3, 1, 5



1 Beskrivende statistikk

For n observasjoner $x_1, \dots, x_n$, og n parvise observasjoner $(x_1, y_1), \dots, (x_n, y_n)$ finner vi:

Gjennomsnitt

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Empirisk varians

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Empirisk standardavvik

$$s = \sqrt{s^2}$$

I dag skal vi...

- ~~Visualisere informasjon om dere og diskutere datatyper (mentimeter)~~
- ~~Regne litt for hånd (oppsummerende tall)~~
- Diskutere populasjon og utvalg (mentimeter)
- Se på sammenhenger (korrelasjon) mellom ulike observasjoner

Fruktbarhetstall

- “Samlet fruktbarhetstall er et mål for hvor mange barn kvinner i en befolkning i snitt vil føde dersom det gjeldende fruktbarhetsmønsteret varer ved.”
https://snl.no/samlet_fruktbarhetstall
- For å finne fruktbarhetstallet kan vi spørre alle norske kvinner (altså hele populasjonen) om hvor mange barn de har.
- Eller, vi kan **estimere** fruktbarhetstallet ved å spørre et tilfeldig utvalg fra populasjonen og regne ut gjennomsnittet.
- Kan vi bruke dere som et tilfeldig utvalg?

Gå til [menti.com](https://www.menti.com)

Populasjon og utvalg

- Fruktbarhetstall i Norge i år 2000: **1.9** (i dag er det 1.5)
- Hvorfor ble vårt tall feil? (diskuter)
- På grunn av utvalget! Vi har ikke data om de kvinnene som ikke har barn, alle vi har data om har minst ett barn
- Utvalget er ikke representativt for populasjonen

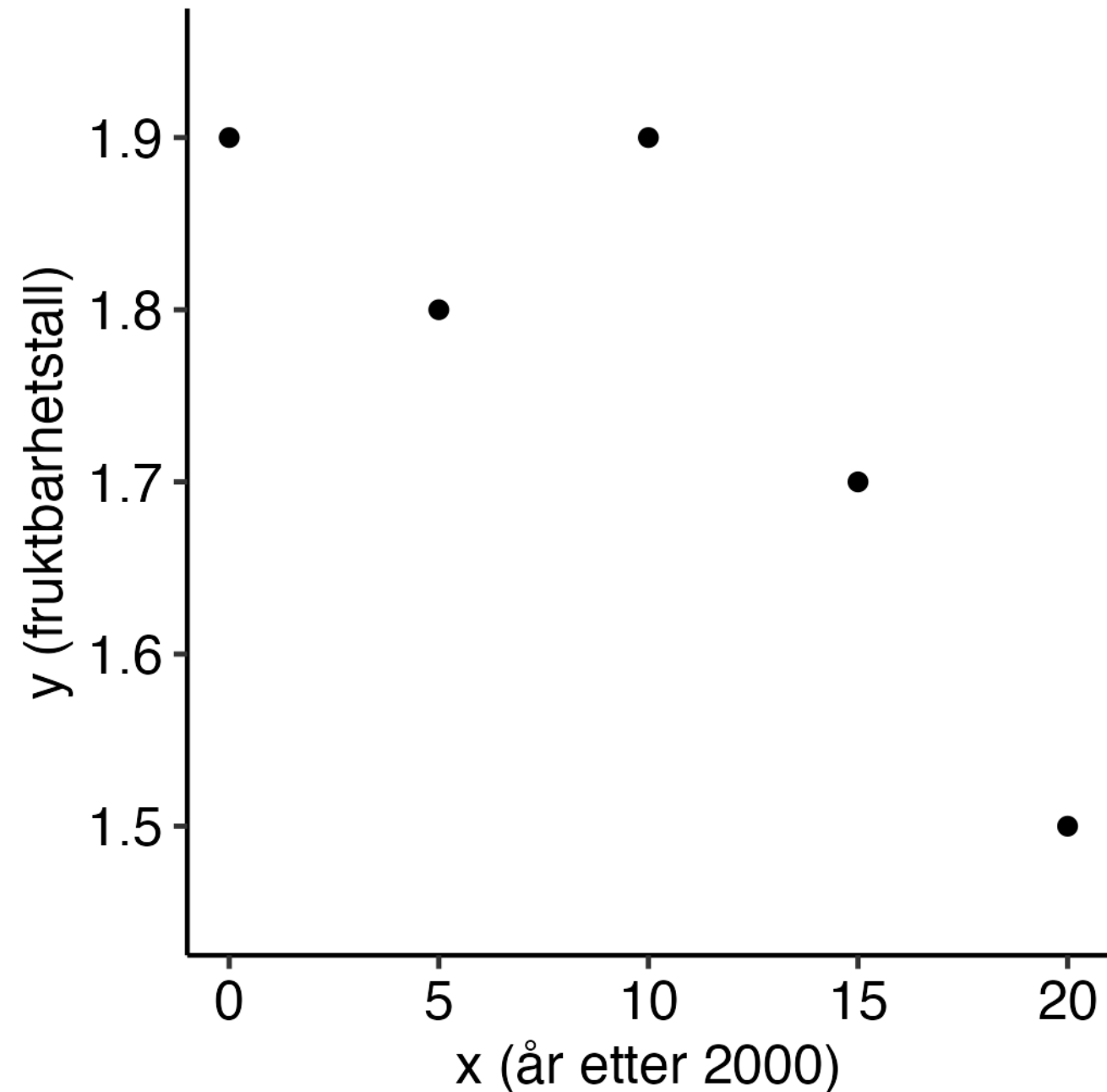
Lineære sammenhenger

År	År etter 2000	Fruktbarhetstall
2000	0	1.9
2005	5	1.8
2010	10	1.9
2015	15	1.7
2020	20	1.5

Kilde: <https://www.fn.no/Statistikk/fruktbarhet>

Empirisk korrelasjon

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



Lineære sammenhenger

År	År etter 2000	Fruktbarhetstall
2000	0	1.9
2005	5	1.8
2010	10	1.9
2015	15	1.7
2020	20	1.5

Kilde: <https://www.fn.no/Statistikk/fruktbarhet>

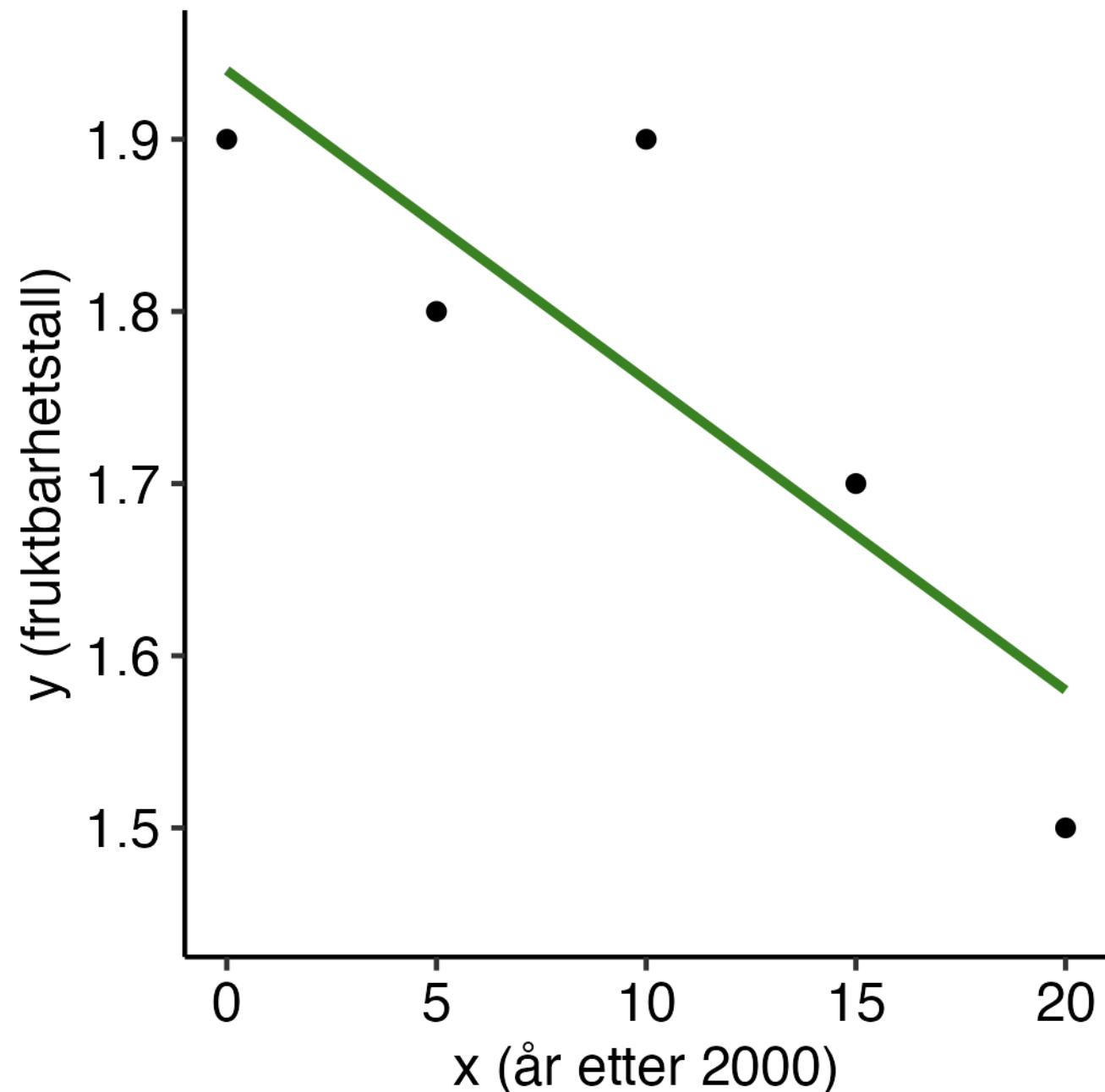
Empirisk korrelasjon

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Minste kvadratsums rette linje $y = a + bx$

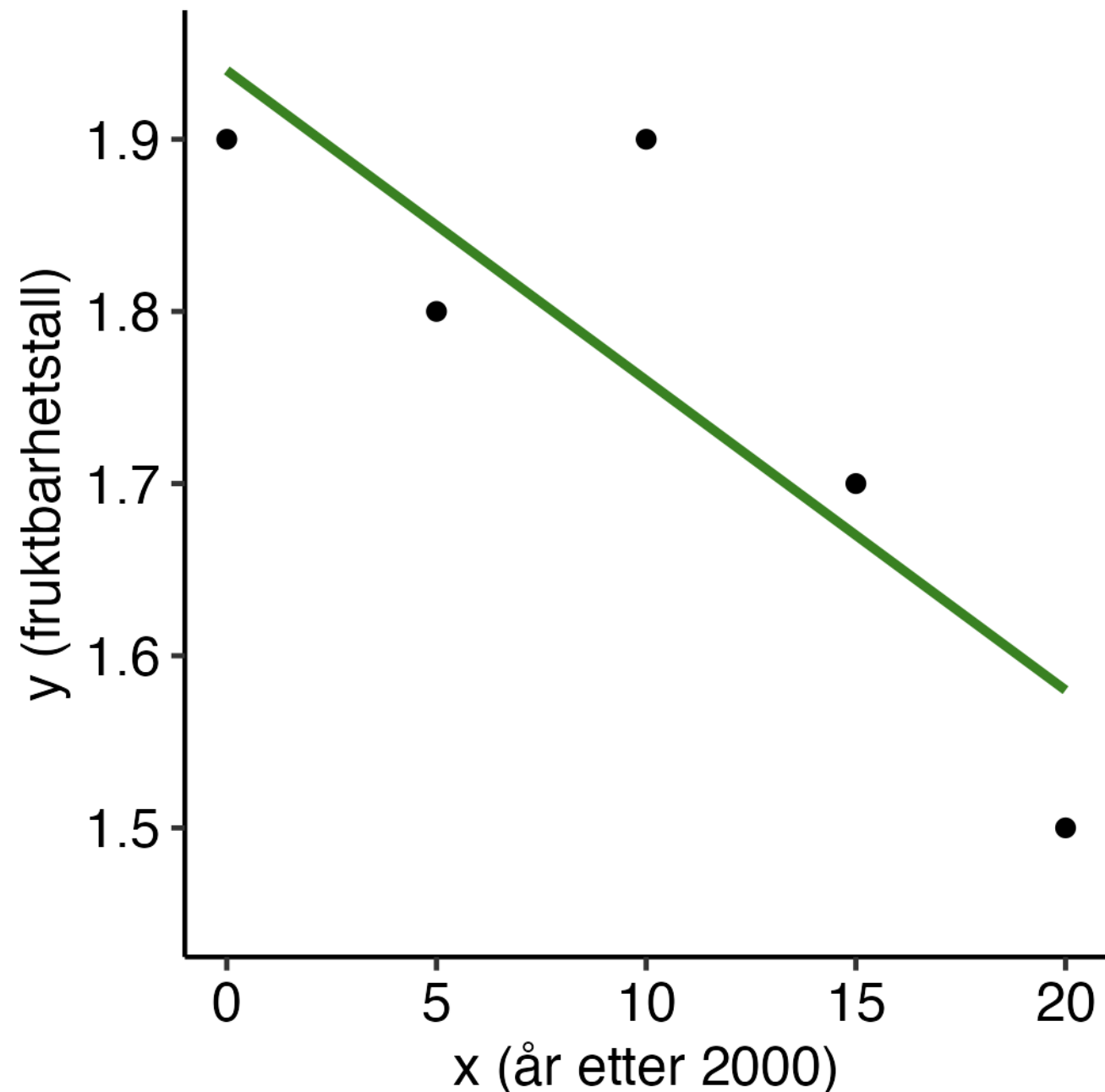
$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}.$$



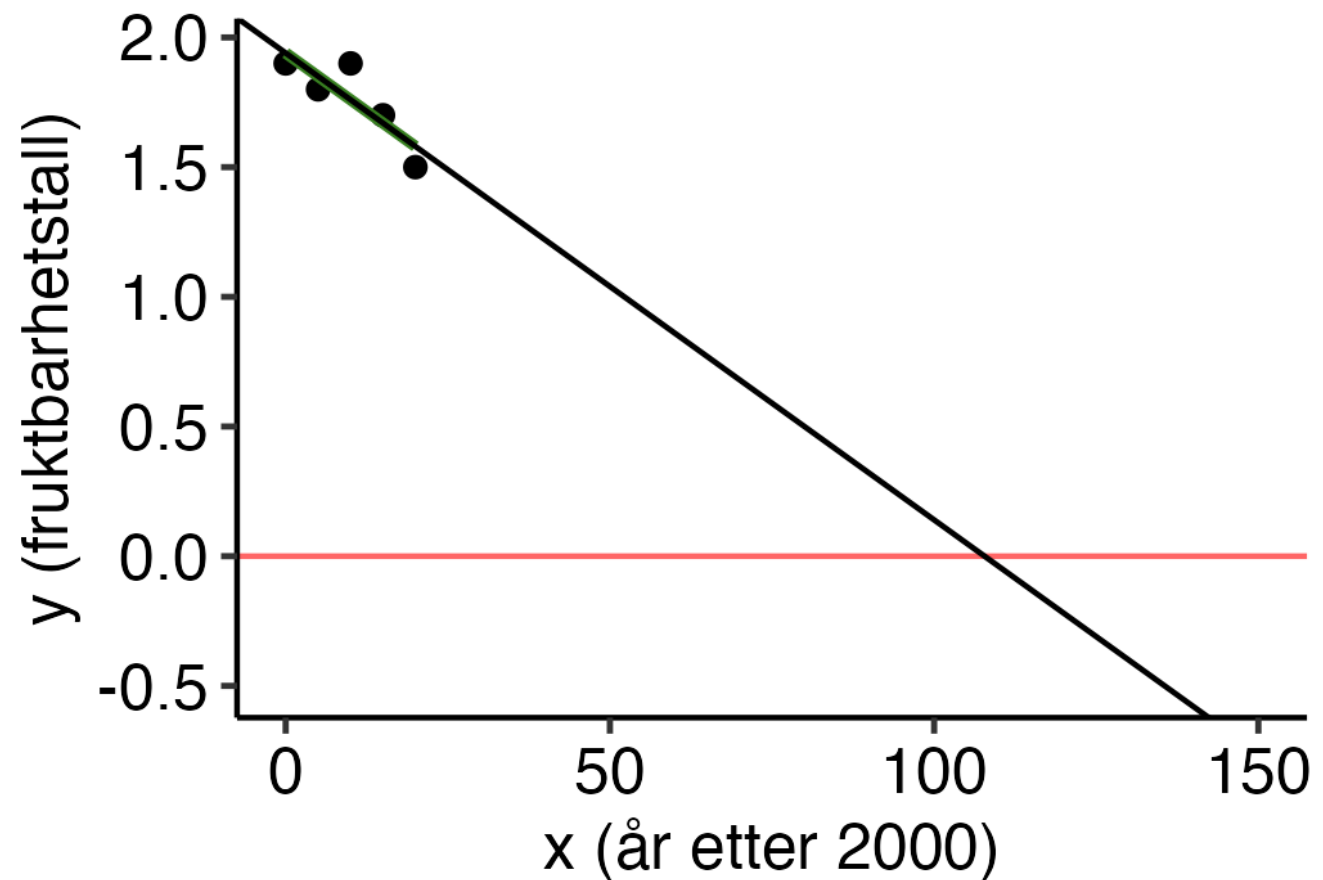
Lineære sammenhenger

- $y = 1.94 - 0.018x$
- “For å opprettholde folketallet på lengre sikt, må fruktbarhetstallet være ca. 2.1” (<https://www.fn.no/Statistikk/fruktbarhet>)
- Dette går jo ikke helt riktig vei...



Lineære sammenhenger

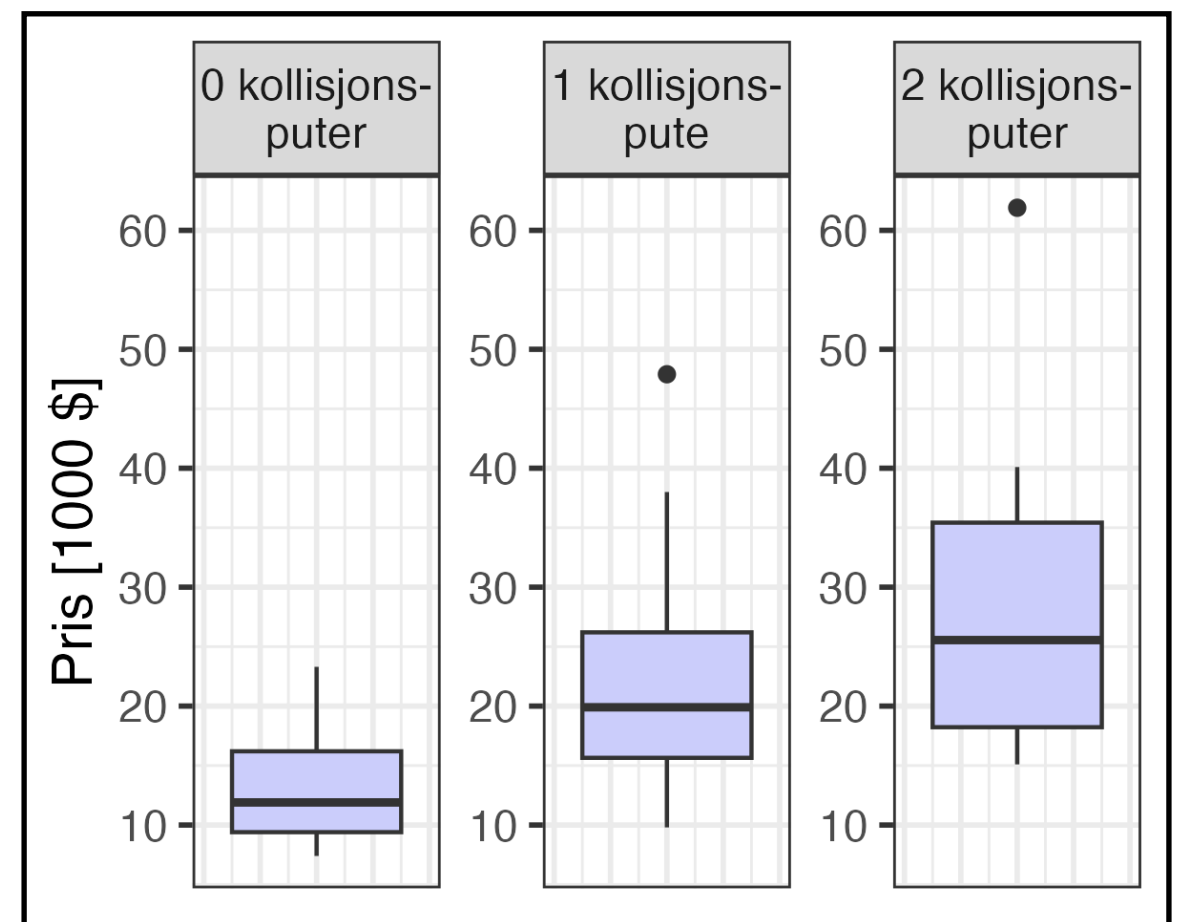
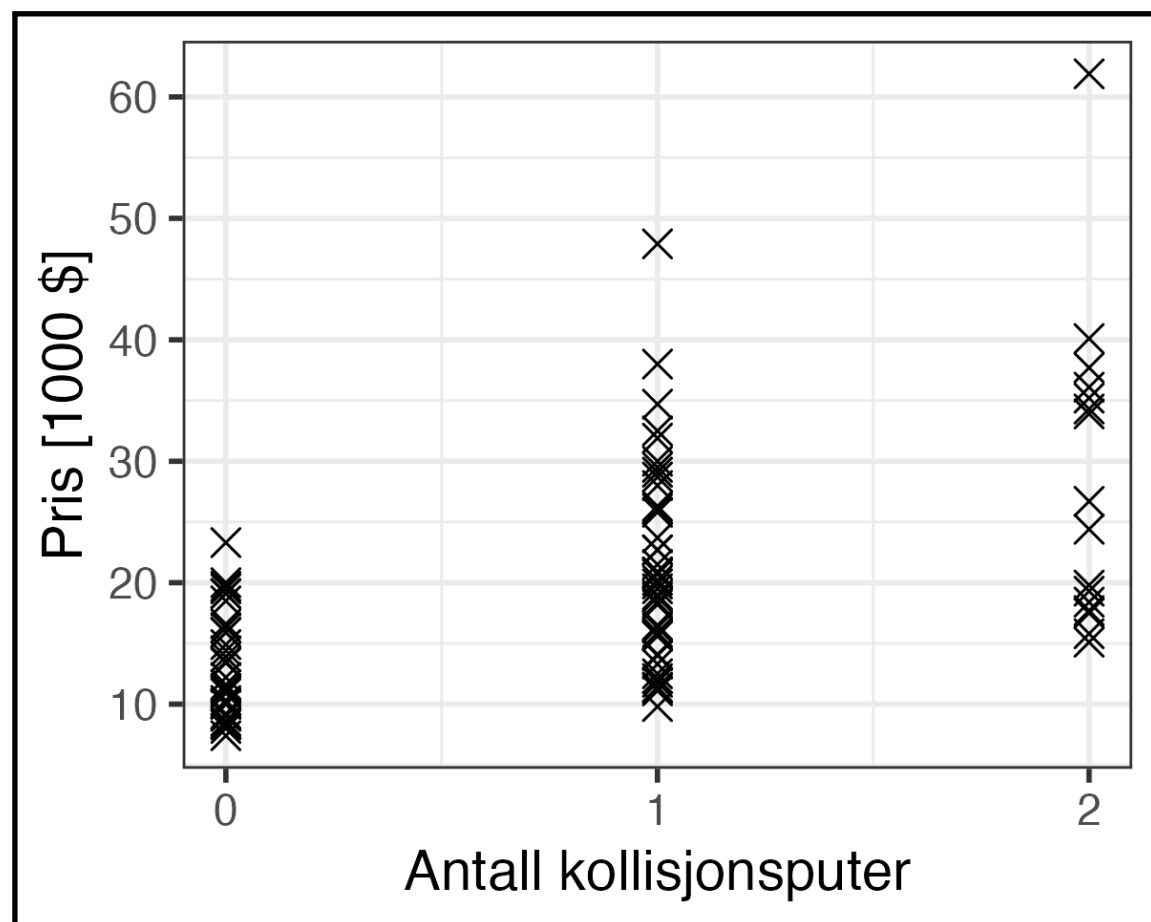
- $y = 1.94 - 0.018x$
- Men vi skal være forsiktige med å anta en fortsatt lineær sammenheng utenfor området med observasjoner!
- Med samme utvikling har vi negativt fruktbarhetstall i 2108
- FN har regnet ut at i 2095 er fruktbarhetstallet i Norge 1.6 (**prediksjon**)



Boksplott

Det åpent tilgjengelige datasettet 93CARS inneholder informasjon om pris, bilmerke, modell, etc. for 93 tilfeldige valgte biler som ble produsert og solgt i USA i år 1993.

Bilmodellene i datasettet har enten 0, 1 (bare til fører), eller 2 (fører og passasjer i framsete) kollisjonsputer. Vi skal studere sammenhengen mellom pris (i 1000 US dollar) og antall kollisjonsputer.



Husk!

- Vil du være med i referansegruppa? Send mail til thea.bjornland@ntnu.no
- Frist øving 1: fredag 23. august klokka 23:59
- Øvingstime torsdag 14:15-16:00 i S5
(utvider til S6 ved behov)