



**UiO** : **Department of Physics**  
University of Oslo

## **HON2200 – Data-drevne prosjekter**

Introduksjon

Oslo, mandag 15. januar 2024

**Eirik Gramstad**

**January 16, 2024**

# Hvem er vi?

## Eirik Gramstad (hovedansvarlig)

- forsker i partikkelfysikk
- jobber med analyse av data fra Large Hadron Collider på CERN

## Halvor Tyseng (gruppelærer)

- tidligere Honours student
- kan introdusere seg selv :-)

## Aksel Braanen Sterri (etikk)

- post-doktor ved OsloMet
- PhD i filosofi og forsker på etiske spørsmål, spesielt de som ligger i skjæringsfeltet mellom helse, teknologi og marked



# Hva skal dere lære?

## Del 1: Innføring i statistikk og databehandling

- forstå og anvende sentrale konsepter innenfor statistikk
  - lineær og logistisk regresjon
  - beslutningstrær
- forstå, bearbeide, analysere og visualisere datasett av moderat størrelse på en informert måte, samt trekke etterprøvbare konklusjoner.
  - [panda dataframes](#) til å lese, behandle og plote data
  - [scikit-learn](#) for statistiske metoder og maskinlæring
  - gjøre hypotesetesting

## Del 2: Etikk i møte med algoritmer basert på input data

- vurdere etiske utfordringer med bruk av ulike typer KI-algoritmer på en systematisk og stringent måte
- se hvordan opprinnelsen til et datasett legger føringer og begrensninger på hva man kan bruke datasettet til

## Del 3: Prosjektoppgave

- jobbe med prosjekter i et tverrfaglig team og kommunisere arbeidet og konklusjonene skriftlig/digitalt og muntlig.

# Hva skal dere lære når?

## Del 1: Innføring i statistikk og databehandling

- **Tidsperiode:** januar (EG)
- i tillegg vil vi arrangere hjelpetimer i forbindelse med innlevering av første obligatoriske oppgave (HT)

## Del 2: Etikk i møte med algoritmer basert på input data

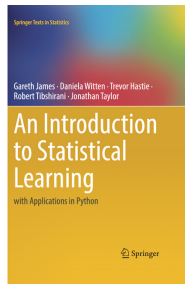
- **Tidsperiode:** februar (ABS)

## Del 3: Prosjektoppgave og relaterte tema

- **Tidsperiode:** mars - mai (EG)
- vil arrangere hjelpetimer ifm. prosjektoppgaven (HT)
- levering av første utkast rundt påske (slutten av mars)
- eksamen vil være i slutten av mai/begynnelsen av juni

# Resurser

- **Pensum:** [An Introduction to Statistical Learning](#)
- **Forelesningsnotater, oppgaver, eksempler, obliker, prosjekter etc.:** [github-side](#)
  - vil bli oppdatert fortløpende
- **Timeplan:** [semestersidene på UiO](#)
- **Innleveringer:** på [UiO-Canvas](#)
- **Diskusjoner/spørsmål:** bruk [Discourse](#)
  - for henvendelser av privat art (sykdom, utsettelse) send en e-post til meg.



Forventer at dere følger med på [timeplanen](#) og gjør evt. forberedelser til hver forelesning

# Vurdering/Eksamen

- én obligatorisk oppgave
  - må være godkjent for å kunne levere endelig prosjekt
  - innlevering midten/slutten av februar
- et prosjekt
  - gruppeoppgave med grupper på tvers av fagfelt
  - levering halveis i prosessen for tilbakemelding
  - endelig prosjektrapport leveres og presenteres i plenum (bestått/ikke bestått)

# Diverse

- Vi trenger to (frivillige) studenter som kan
  - ta imot meldinger fra medstudenter
  - si ifra om ting ikke fungerer optimalt
  - delta på emnets avsluttende evalueringsmøte og formidle ris og ros.

Frivillige?

**SPØRSMÅL?**



# La oss sett i gang med det som er gøy!

- erfaringsmessig er det å kunne behandle store mengder data, effektivt analysere de og trekke fornuftige konklusjoner fra de veldig etterspurt i næringslivet

# Predikasjon

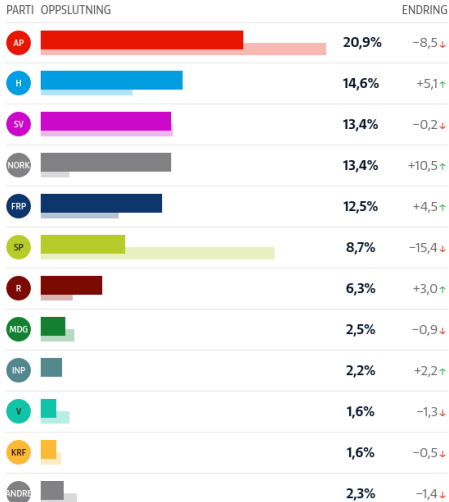


Figure: Meningsmåling (uken før valget)

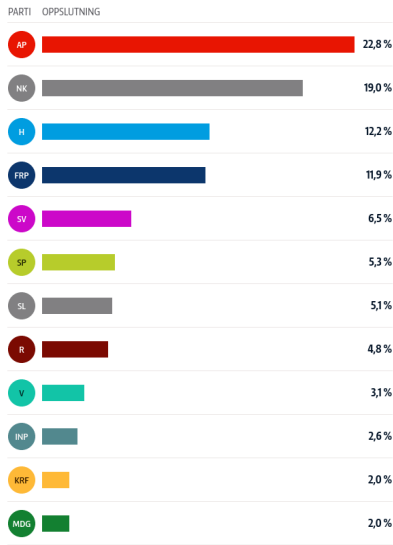


Figure: Valgresultat

# Statistisk signifikans: 800 forskere advarer mot misbruk av metode

Forskere bak opprop krever endring.

**Eksempel:** Er det en sammenheng mellom bruk av en spesifikk type NSAIDs og hjerterytmeforstyrrelse?

Schmidt et al.	YES
Chao et al.	NO

# Statistisk signifikans og p-verdi

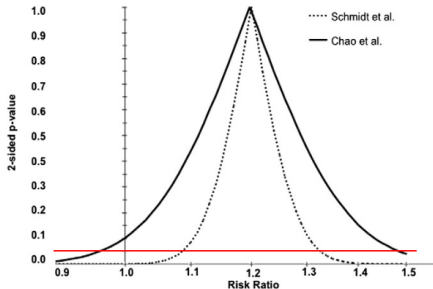


Figure: Ref.: [Schmidt, M. and Rothman, K. J. Int. J. Cardiol. 177, 1089–1090 \(2014\)](#)

# Statistisk signifikans og p-verdi

Et **ikke-signifikant** resultat beviser ikke at *null-hypotesen* er korrekt.  
Et **signifikant** resultat beviser heller ikke noen annen hypotese.

# Statistisk signifikans og n-verdi

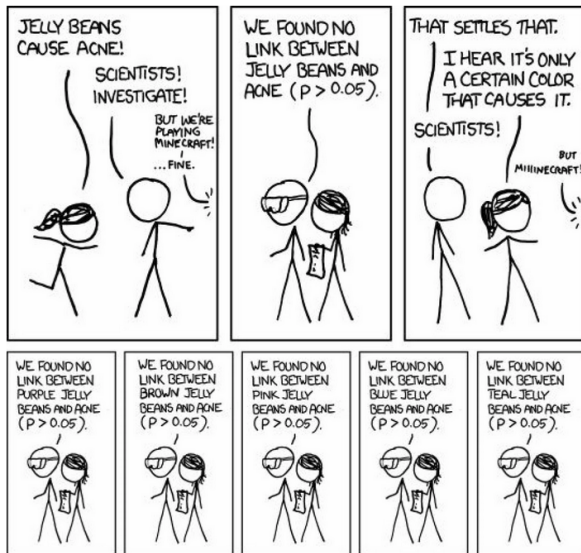


Figure: 1/3

# Statistisk signifikans og p-verdi

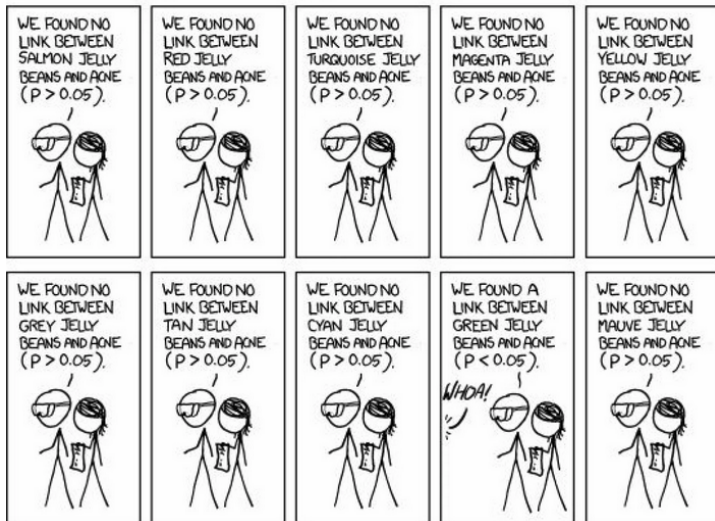


Figure: 2/3

# Statistisk signifikans og p-verdi

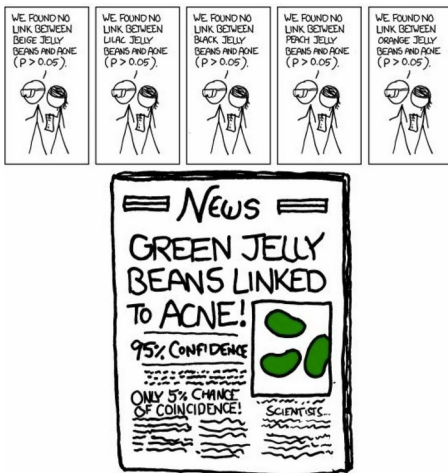


Figure: 3/3



# Statistisk signifikans og p-verdi

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	

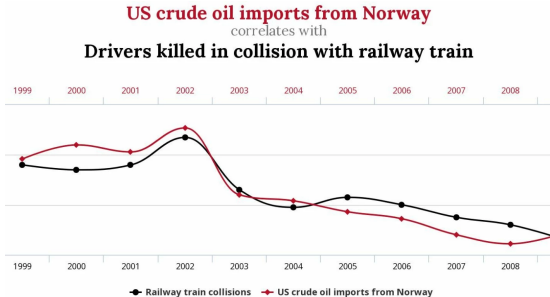


Figure: Ref.: <http://tylervigen.com/spurious-correlations>