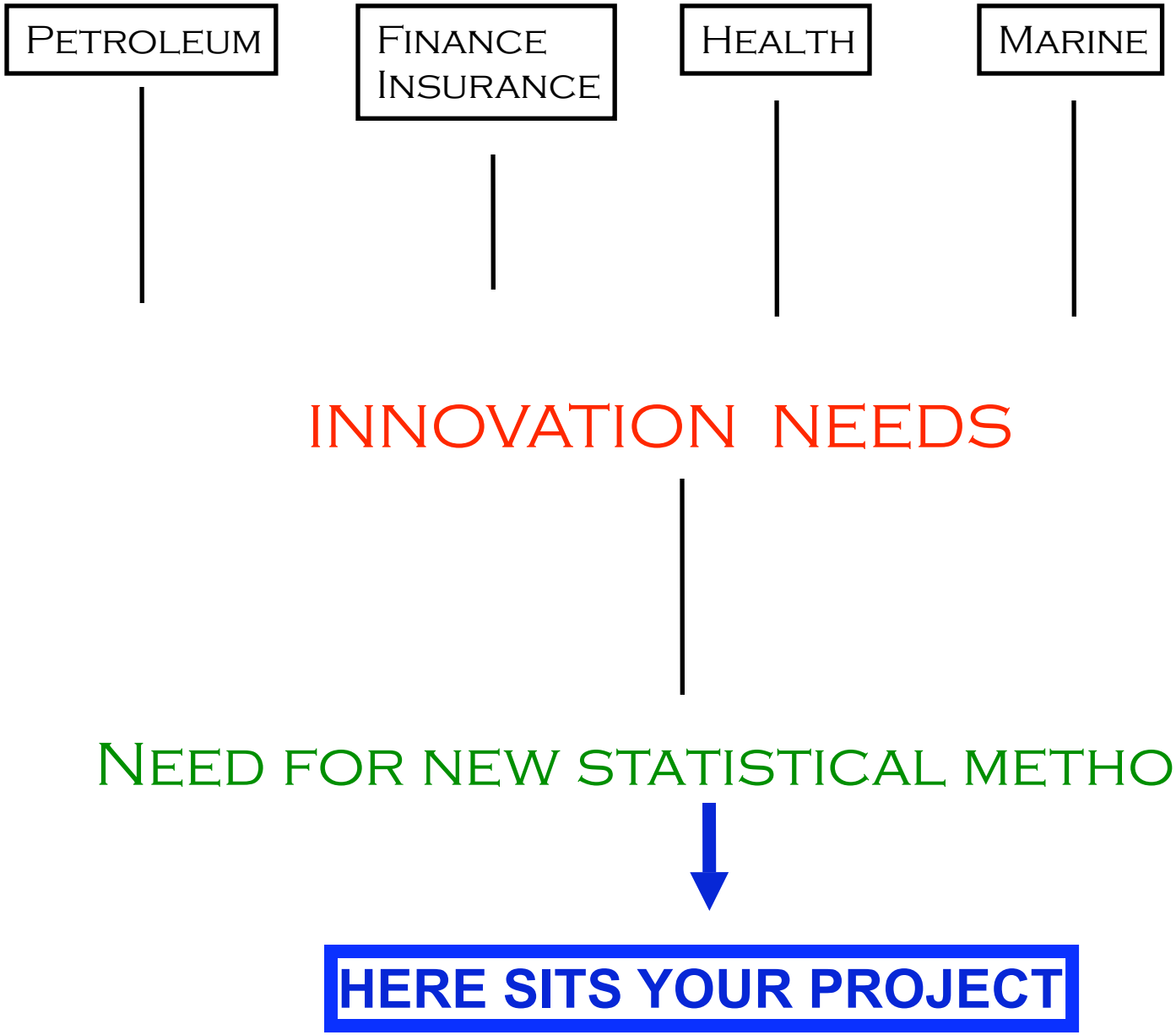
Sentre for forskningsdrevet innovasjon

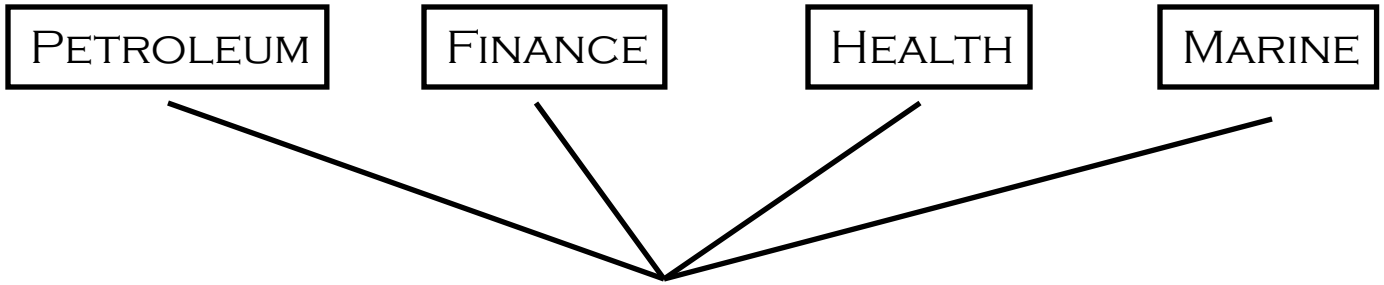
(sfi)² Statistics for Innovation

Arnoldo Frigessi
frigessi@medisin.uio.no

Ørnulf Borgan
borgan@math.uio.no

WITH THE SUPPORT OF Norges forskningsråd


PETROLEUM
FINANCE
INSURANCE
HEALTH
MARINE
INNOVATION NEEDS
NEED FOR NEW STATISTICAL METHODS
HERE SITS YOUR PROJECT



- much data from different sources, (almost too much data)
- more complex data, because we measure more in depth
- necessary to discover “second order effects”
- computationally intensive procedures

HYDRO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

HYDRO

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

HYDRO

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

NTNU
Det skapende universitet

HYDRO

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

NTNU
Det skapende universitet

DnB NOR

HYDRO

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

NTNU
Det skapende universitet

DnB NOR

GJENSIDIGE

HYDRO

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

NTNU
Det skapende universitet

DnB NOR

GJENSIDIGE

HYDRO

STATOIL

UNIVERSITETET
I OSLO

NR
Norsk
Regnesentral
NORWEGIAN COMPUTING CENTER

NTNU
Det skapende universitet

DnB NOR

GJENSIDIGE

HYDRO

STATOIL

HAVFORSKNINGSINSTITUTTET
INSTITUTE OF MARINE RESEARCH

Rikshospitalet - Radiumhospitalet HF

Rikshospitalet - Radiumhospitalet HF

Smerud
Medical Research

Smerud
Medical Research

Multicompare

- Assume you need to monitor, understand, compare, or predict very many dependent things, simultaneously.

  For example:

  - 30000 genes, to find which are causing a disease;
  - 10000 financial products, to detect strange behaviours;
  - 5000 clients, to find the most dangerous ones.

- We have such data and questions in many areas!

- Classical tests fail:
  it produces too many false positives
  and false negatives.

Multicompare

- Recently, an explosion of new and better approaches.
  - False discovery rate
  - Multiple testing
  - Sparsity …
- We have developed a new method and code, that takes advantage of multiple data sets (data integration)
- We have applied it to genomics.
- What shall you do?
  - understand the problem
  - understand the new method
  - apply it to a new area, for example in finance and/or
  - extend the method, so that it is more flexible to the type of test

MicroSurv

- Regression with many covariates and much less observations:

  - 10 000 covariates (x_1, … , x_10000)
  - 200 cases (y)

- p>> n

- Impossible to do usual (logistic, Cox) regression

- Methods to reduce the number of covariates, finding the important ones (lasso, ridge regression, variable selection …)

- We have done cases in genomics and want to do other type of data now.

- Macro-economics? Biology? Finance?

Arnoldo Frigessi
frigessi@medisin.uio.no