

Hva er statistikk, og hva gjør en statistiker?

Mette Langaas

Institutt for matematiske fag, NTNU, og Norsk Regnesentral

21. februar 2017

Høyde blant rekrutter

108 Egenrapportert høyde og vekt for sesjonspliktige¹

	Gjennomsnittshøyde, cm		Gjennomsnittsvekt, kg	
	Gutter	Jenter	Gutter	Jenter
1910	171,0	..	..	..
1920	171,4	..	..	..
1930	172,8	..	..	..
1937	173,8	..	..	..
1952	176,2	..	..	..
1960	177,1	..	..	..
1970	178,7	..	..	..
1980	179,4	..	..	..
1990	179,7	..	..	..
2000	179,9	..	72,8	..
2005	179,8	..	74,7	..
2010	179,5	..	75,5	..
2011 (født 1994)	179,9	167,1	73,7	62,1
2012 (født 1995)	179,9	167,1	73,9	62,5

¹ Brudd i statistikken i 2011 pga. endring i Vernepliktsloven fra 1. januar 2010. Tidligere årganger har vært basert på fysiske målinger ved sesjon. Fra 2011 benyttes opplysninger fra egenrapportert nettskjema som sendes Vernepliktsverket normalt det året de fyller 17 år.

Kilde: Vernepliktsverket.

Mer informasjon: <http://www.forsvaret.no/>

<http://www.ssb.no/a/aarbok/tab/tab-108.html>

Hva er statistikk?

- ▶ Statistikk har som mål å utvikle vår kunnskap basert på innsamling og analyse av empiriske data.
- ▶ To grener:
 - ▶ Sannsynlighetsteori: matematisk studium av sannsynlighet for tilfeldige hendelser.
 - ▶ Statistisk inferens: modeller og metoder for å samle inn, beskrive, analysere og fortolke numeriske data.



Grunnkurs i statistikk ved NTNU, IMF: ST0103, ST1101, ST1201, TMA4240, TMA4245.

Den vitenskapelige prosess

[1] veldefinert forskningsspørsmål

Den vitenskapelige prosess

[1] veldefinert forskningsspørsmål



[2] designe studie – samle inn data

Den vitenskapelige prosess

[1] veldefinert forskningsspørsmål

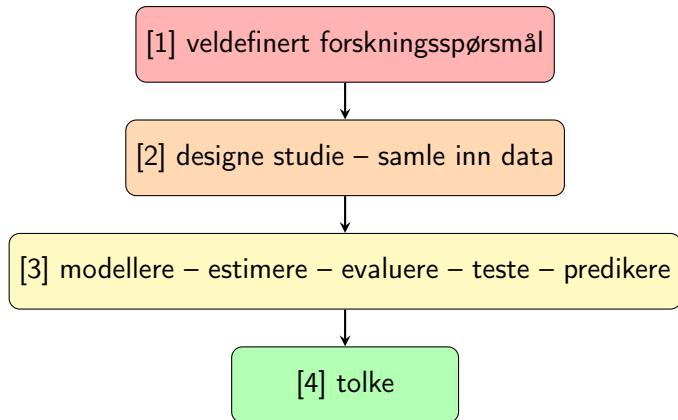


[2] designe studie – samle inn data

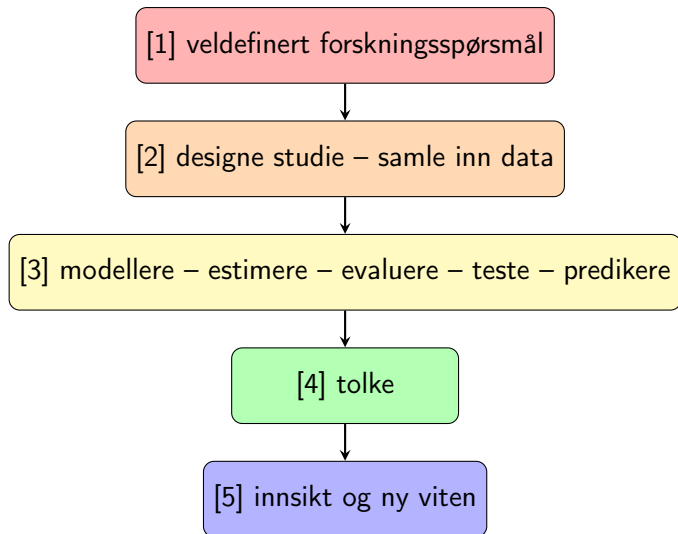


[3] modellere – estimere – evaluere – teste – predikere

Den vitenskapelige prosess



Den vitenskapelige prosess



Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

Jeg har jobbet/jobber med:

1. Hvor mange vågehval er det i nordøst-Atlanteren?
2. Hvor mange aviser skal Dagbladet levere til hvert utsalgsted?
3. Hvilke gener hos våskrinneblomst er aktive etter at bladene på planten er blitt skadet?
4. Hvordan kan vi identifisere genetiske markører som er assosiert med sykdom når det er viktig ikke gjøre falske positive funn?
5. Hvis vi har begrensede midler, vil det da lønne seg å samle inn genetiske data fra ekstreme individer i en populasjon istedenfor å gjøre et tilfeldig utvalg?
6. Hvordan vokser en kreftsvulst? Kan vi teste ulike modeller fra data på mutasjonsfrekvenser i en svulst?
7. Hvordan kan vi lage en modell for hvordan nervecellene i motor cortex hos mus jobber når musene utfører taktiske oppgaver?
8. Hvordan kan man raskt oppdage avvik fra normal drift fra sensor-data fra skip?

The best thing about being a statistician is that you get to play in everyone's backyard (*John Tukey*).

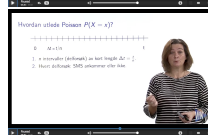
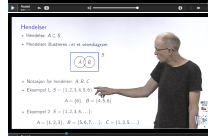
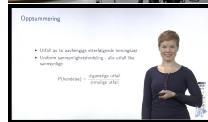


- ▶ Siv.ing fra FysMat og studieretningen for Industriell matematikk (NTH) i 1991.
- ▶ Dr. ing. i matematiske fag (statistikk), NTH, fra 1995.
- ▶ Forsker ved Norsk Regnesentral, Oslo, 1995-2001.
- ▶ Postdoc ved Institutt for matematiske fag, NTNU, 2002-2004.
- ▶ F.aman i statistikk ved Institutt for matematiske fag fra 2004, Professor fra 2014.
- ▶ F.aman. ved Enhet for anvendt klinisk forskning, Fakultetet for medisin, 2009-2010.
- ▶ Forskningsterminer: Melbourne 2008-2009 og Perth 2014-2015, Australia.
- ▶ Bistillinger (20%) ved Universitet i Oslo, og nå ved Norsk Regnesentral.



Kvalitet, tilgjengelighet og differensiering i grunnutdanningen i matematikk

- ▶ Innovativt undervisningsprosjekt med hovedmål å øke læringsutbyttet for studentene.
- ▶ TMA4100 Matematikk 1, TMA4105 Matematikk 2 og TMA4240/TMA4245 Statistikk.
- ▶ Mattelab og digitale øvinger: Tilgjengelige faglærere og rask respons på egen innsats fremmer læring.
- ▶ Aktiv læring i forelesninger: Matematikk er ingen tilskuersport, men studenter kan aktiveres, også i store auditorier.
- ▶ Digitale ressurser: De digitale ressursene er laget for studenter på ulikt faglig nivå, og er tilgjengelige hvor som helst og når som helst.



Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Nyttig hvis man skal jobbe innen statistikk/data-analyse

- ▶ Ha kunnskap om statistisk tankemåte, statistiske modeller og metoder - vi har mange spennende fag på NTNU!
- ▶ Programmeringskunnskap (Python, C++, statistisk software).
- ▶ God kunnskaper innenfor et spesifikt (anvendelses)-område.
- ▶ Synes det er artig å (forsøke å) forstå (deler av) verden ved å lage modeller, analysere data og trekke konklusjoner - være strukturert og kunne jobbe både teoretisk og praktisk.
- ▶ Kunne kommunisere med "den som har problemet" - samarbeid og formidling.
- ▶ "Complex problem solving skills" - for fremtiden.

Prediksjon: i fremtiden (neste 5-10 år) vil etterspørselen etter "data-analyse-kunnskaper" øke

[HR TODAY](#)[RESOURCES & TOOLS](#)[LEARNING & CAREER](#)[EVENTS](#)[ENTIRE SITE](#)[RESEARCH & SURVEYS](#)

Jobs of the Future: Data Analysis Skills

Nov 14, 2016

[LIKE](#)[SAVE](#)[PRINT](#)[REUSE PERMISSIONS](#)

The purpose of this research is to inform workforce planning efforts for jobs of the future—occupations that are predicted to grow over the next 5-10 years. The report, sponsored by the American Statistical Association, covers the supply and demand for data analysis skills, the function and types of employees needed for these jobs, and skill and education requirements at different levels.

[DOWNLOAD FULL SURVEY FINDINGS \(PDF\)](#)

Key findings include:

- 59% of organizations expect to increase the number of positions requiring data analysis skills over the next five years.
- 78% of organizations who had hired for data analysis positions in the last 12 months reported difficulty recruiting for these positions.
- The most common functional areas for data analysis positions are accounting and finance (71%), human resources (54%), and business and administration (50%).

Noen ressurser

- ▶ Statistiker: <http://studenttorget.no/statistiker>
- ▶ Emner ved IMF: <https://wiki.math.ntnu.no/emner>.
- ▶ Etter grunnkurs: TMA4265 Stokastiske prosesser, TMA4255 Anvendt statistikk eller TMA4267 Lineære statistiske modeller, og fra våren 2018 TMA4268 Statistisk læring.
- ▶ Mer om KTDiM: www.ntnu.no/ktdim og wiki.math.ntnu.no/ktdim (med lenke til ressurser i Matte 1, 2 og Statistikk).
- ▶ Statistrikk-bloggen: www.statistrikk.no