

Plan for dagen

- Kap 1: Deskriptiv statistikk, populasjon, utval
- Kap 2: Sannsynsteori
- Velje referansegruppe
- Kahoot

Heimeside:

<https://wiki.math.ntnu.no/tma4240/2016h/start>

Deskriptiv statistikk

og kalla beskrivande statistikk. Finn ut oppsummerande storleikar, d.v.s. tal, eller visualiserer datasett ved ulike plott.

OPPSUMMERANDE STORLEIKAR for datasett

$$x = \{x_1, x_2, \dots, x_n\}$$

Lokasjon:

- Gjennomsnitt (mean), $\bar{x}$
- Median: Midterste datapunktet når dei er sortert i stigande rekkjefølgje.
- Typetal (mode): Dataverdien det er flest av

Spredning

- Empirisk varians: $Var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- Empirisk standardavvik: $Std(x) = \sqrt{Var(x)}$
- α -kvantil: Ein andel α av dataene er mindre eller lik denne verdien.
- Kvartilbredde: 0.75-kvantil - 0.25-kvantil

- Finne relevante observatorar og nokre plott er alltid ein god start, og kan spare deg for mykje arbeid seinare.
- Det gode plottet kan fortelje meir enn 1000 ord.
- Ein god figur krever akseinfo/einingar og forklaringar. Ja, det tar tid! Sjå notata under *Matlab filer og eksempler* på <https://wiki.math.ntnu.no/tma4240/2016h/matlab> for inspirasjon. og video frå Hans Rosling <http://www.gapminder.org/videos/dont-panic-the-facts-about-population/>

Snøtetthet: Treng vi mål snøtetthet for å finn SWE

(Meir info, data og matlab-kode på

<https://wiki.math.ntnu.no/tma4240/2016h/matlab>)

Bakgrunn

Snømengde er viktig for å planlegge / prise for vasskraftselskap, og for å forutsjå (vår)flom. Billig å måle snødybde, dyrt å måle snøtetthet.

Har for to stadar i Sylan (Rybekken og Sylsjødammen) målt snødybde og snøtetthet over fleire år. I fila:

- År
- Snødybde (cm)
- Snøtetthet (gr/cm^3)
- SWE (snow water equivalent) = snødybde $\times$ snøtetthet (cm)
- DOY (day of year), 32 = 1.februar.

Kva er statistikk?

Statistikk er vitskapen om læring frå data, og måling, kontroll og kommunikasjon av usikkerheit (Davians Louis, Science, 2012).

Statistikk er eit verktøy for å gjere vurderingar og ta beslutningar under usikkerheit og variasjon.

Lære frå data: Vil ofte trekke slutningar utover det observerte. Vi vil seie noko om *populasjonen* frå eit *utval*.

- Bør laks røykast med metode 1 eller metode 2?
- Kva er gjennomsnittshøgde for unge norske kvinner?
- Er NTNU studentar høgare enn den unge norske befolkningen?
- Gitt høgda på foreldra, kor høgt blir barnet?
- Kor stor andel av nordmenn er blåøyd og blond?
- Er det nok å måle snødybde for å finne vassekvivalent (SWE)?
- Ønske?

Ein **populasjon** i statistikk er ei samling av objekt eller hendingar.

- Kan vere *konkret* (f.eks. alle barn født i 2016 eller alle ventiler av en bestemt type) eller mer *abstrakt* (f.eks. alle barn 'som kunne blitt' født i 2016 eller alle ventiler av denne typen 'som kunne ha blitt produsert')
- Oftest abstrakt, knyttet til en statistisk modell.

Vi lærer frå data (eit utval) ved å spesifisere ein statistisk modell.

- **Statistisk modell:** Ein (idealisert) modell for korleis data har oppstått.

Vi bruker sannsynsteori for å spesifisere en statistisk modell.

Sannsynsteori

Den delen av matematikken som omhandlar stokastiske (tilfeldige) fenomen. Matematisk abstraksjon av ikkje-deterministiske fenomen.

Kva utvalg kan vi bruke:

- Spelarar på NTNUI-basketball for å seie noko om høgde på NTNU studentar?
- Oss for å seie noko om høgde NTNU-studentar?
- Oss til å seie noko om samanheng høgde foreldre-barn?
- Flyplasstilfredshet frå dei som trykkar på smilefjes?

Samlar inn data, og vi vil trekke slutningar utover det observerte.
OK dersom vi har eit representativt utval.

- Løgn, forbanna løgn og statistikk
- Om ein torturerer dataene sine lenge nok, vil dei tilstå.

Når ein samlar inn, bearbeider, analyserer og tolker data gjer ein mange val / antakingar.

Ver ærleg med deg sjølv.

Ellers torturerer du dataene, og driv på med noko verre enn løgn.

...hugs referansegruppe....