

Statistik och Dataanalys I

Föreläsning 1 - Introduktion - EXPERIMENT!!!

Mattias Villani



Statistiska institutionen
Stockholms universitet



mattiasvillani.com



@matvil



[@matvil](https://mattiasvillani.com)

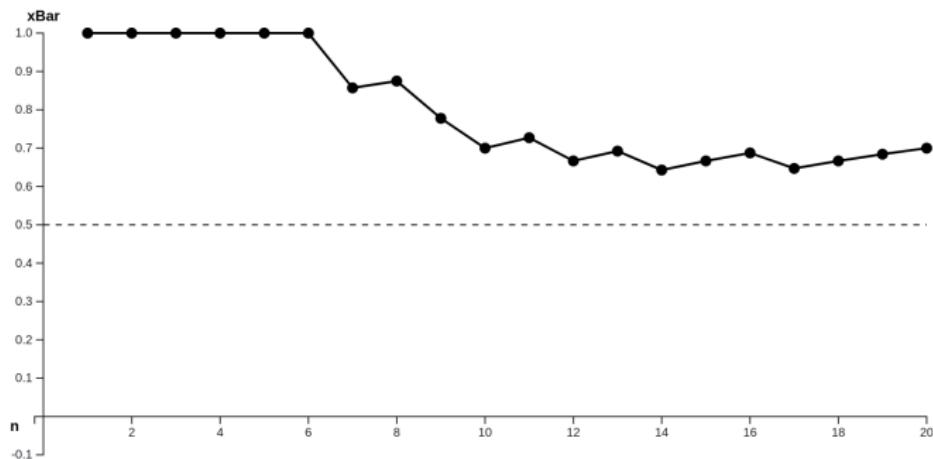
Översikt

- Motiverande exempel
- This
- That

Stora talens lag



$$p \approx \frac{14}{20} = 0.7000$$



```
setup = undefined
```

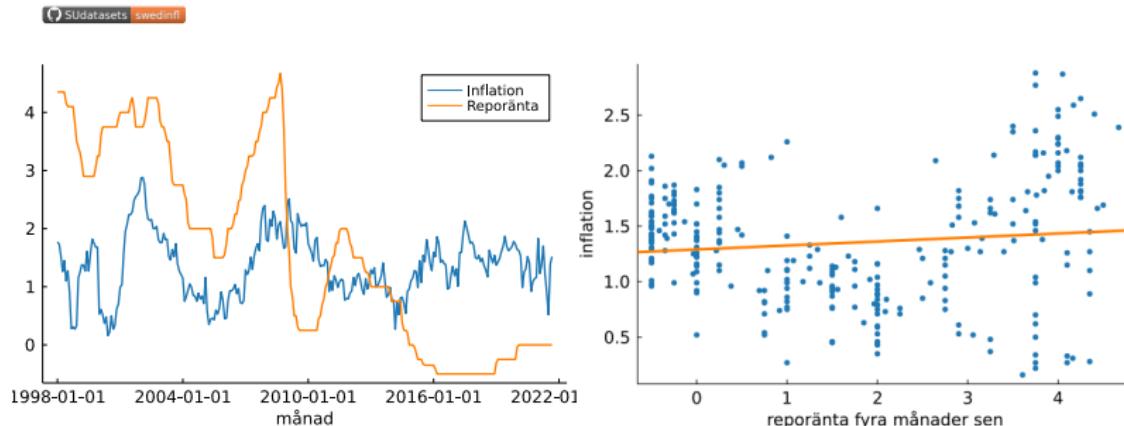
 Mattias Villani Stora talens lag

 Observable

Adaption från Marc McClures [Law of Large Numbers](#)

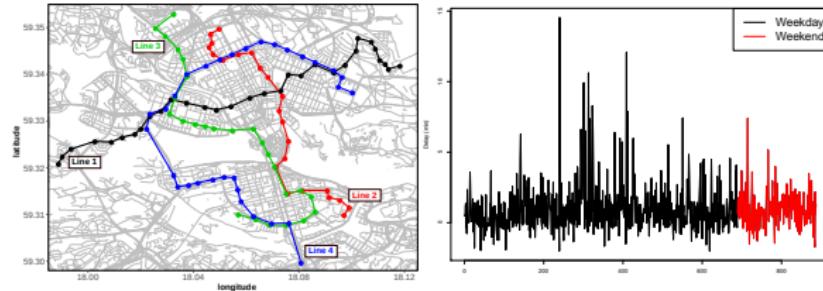
Riksbanken och styrräntan

- Riksbankens mål är att hålla inflationen nära 2% per år.
- Riksbanken bestämmer den s k reporäntan in ekonomin.
- Hur beror inflationen på räntan?
- Både inflation och reporänta är exempel på **tidsserier**.
- **Sambandet** mellan inflation och ränta: **regression**.



Förseningar i lokaltrafiken

- Mål1: **förutsäga förseningar** för stadsbussar.
- Mål2: **säkerheten** i prediktionen: **5 min, 5 min, 5 min**
- Data: alla förseningar för alla busslinjer i Sthlm under 1 år.
- Mål: förutsäga förseningen för 12.15-bussen till Tegnérsgatan.
- Förklarande variabler:
 - ▶ försening för 12.15-bussen vid hållplatser innan Tegnérsgatan.
 - ▶ förseningar för tidigare bussar vid hållplats Tegnérsgatan.
 - ▶ tid på dagen
 - ▶ rusningstid?

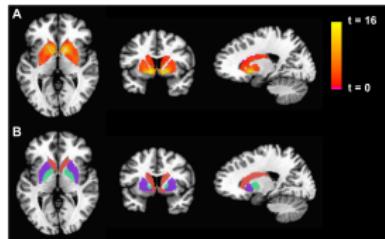


Var i hjärnan skapas vårt språk?

- Person i MR scanner pratar omväxlande med att knyta handen.

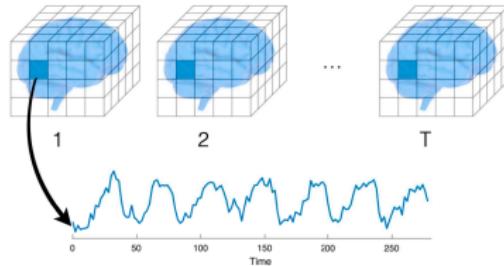


Lars Kruse, AU Kommunikation, CC license



[Source](#), CC license

- Mäter mängden syresatt blod på tusentals ställen i hjärnan (data).



Artificiell intelligens och maskininlärning

- Mål: få en maskin att känna igen handskrivna siffor.
- Data: 60000 handskrivna siffror mellan 0-9.
- Varje bild har 28×28 pixlar med värde mellan 0 och 255:

0 = svart

128 = mellangrå

255 =



- **Binär regression**: modell för *sannolikheten* för 1:a.
- Djupa neural nätverk (deep learning) är en form av **regression**.

Lära en maskin att känna igen siffror

- Mål: få en maskin att känna igen handskrivna siffror.
- **Bilder är data:**
 - ▶ 60000 handskrivna siffror mellan 0-9.
 - ▶ Varje bild har 28×28 pixlar med värde mellan 0 och 255:

$0 =$ svart

$128 =$ mellangrå

$255 =$



- **Binär regression:** modell för *sannolikheten* för 1:a.
- Djupa neural nätverk (deep learning) är en form av **regression**.

Artificiell intelligens och maskininlärning



- Statistik är grunden för modern AI.



... the reader should have some knowledge of basic statistics, including variance, correlation, simple linear regression, and basic hypothesis testing (e.g. p-values and test statistics).

- Deep Learning Book: Kapitel 3:
- Sannolikheter, slumpvariabler, sannolikhetsfördelningar, väntevärde, varians, kovarians, korrelation, regression, Bayes sats, Bernoulli, exponential, Normal, osv.

Statistik ett framtidsyrke

Statistiker får jobb som data scientists