**Lineær regresjon**

Når vi bruker linær regresjon ønsker vi å kunne predikere, altså forutse, hvordan en gruppe oppfører seg. Dette gjøres ved å ta et tilfeldig utvalg av gruppen, og bruke denne til å kunne bestemme egenskapene til hele gruppen.

En regresjonsmodell er lineær når alle termene i likningen/modellen er en konstant eller en parameter multiplisert med en independent variabel.

Lineære modeller kan modellere kurvatur, ved at man inkluderer ikke-lineære variabler slik som for eksempel polynomer og transformere eksponentielle funksjoner.

**Ordinary least squares**

<https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/>

OLS er den vanligste formen for lineær regresjon.

Så lenge modellen oppfyller OLS forutsetningene for lineær regresjon, så vil denne metoden gi de beste estimatene. Dette betyr at de estimerte verdiene ligger veldig nærme de faktiske verdiene, som vil si at vi har lav varians.

La oss se på hvilke forutsetninger for lineær regresjon som må være oppfylt for at OLS skal gi de beste estimatene. Flere av disse forutsetningene handler om feilen (error, menes noise her?), som vil være umulig å si noe presist om (?). En ting som er nyttig å se på er residuals. Dette betur å se på differansen mellom den observerte/faktisk verdien, og den fitted verdien.

OLS estimerer betaene.

En annen forutsetning for OLS er at erroren/noisen skal ha en gjennomsnittlig verdi på null. Den randome noisen tvinger gjennomsnittsverdien til residuals til å bli null (??).

Den randome noisen må ikke være normalisert, men det gjør det mulig å få troverdige confidence intervaller og prediksjonsintervaller.

**MSE**

Den gjennomsnittlige kvadratiske differansen mellom den estimerte verdien og den faktiske verdien.

Den angir kvaliteten på estimatet (eller prediksjonen) og er alltid positiv og skal være så nærme null som mulig. Definisjonen avhenger av om man skal beskrive en estimator eller en predictor.

**R2 Score**

<https://en.wikipedia.org/wiki/Coefficient_of_determination>

the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

**K-fold / resampling technique**

…

**Ridge**

…

**Lasso**

…