

Partikkel-Prosess (Partikkel-Prosess (5+9var).00d)

Fremstilling av partikler som skal benyttes som ultralyd-kontrast-middel.

Det betyr at størrelse og størrelses-fordeling er svært viktig.

Settet inneholder et fullfaktor-design for 5 variable + 8 senterpunkter

Det er 9 responsvariable (ulike mål for størrelse og størrelsesintervall)

Målet er å finne en prosess som gir størst mulig utbytte for partikler med en middels størrelse.
(Dvs. mange mellom 5 og 8 (evt. 4 og 10) og samtidig få svært store eller små)

Kan i prinsippet modelleres med MLR, men PLS-1 eller PLS-2 kan kanskje gi noen fordeler(?)

Predikere biologisk aktivitet (Predicting Biological Activity.xlsx)

The column **obsnam** contains an **identification code**.

Each record represents a peptide chain of five amino acids (AA). Each amino acid name is coded using a single letter and each chain is represented by five letters, as shown in the column obsnam.

The descriptors used for modelling the response are **size**, **hydrophobicity**, and **polarity** of the AA at the various sites in the molecule.

The response of interest is **rai**, a relative measure of Bradykinin potentiating activity.
However, rai is highly skewed, and so **log_rai**, the base 10 logarithm of rai may be used.

The study was blocked.

The first 15 samples were studied first, and then, at a later time, the next 15 samples.

The choice of AA-sequence does not seem to be based on any experimental design.
Is this a correct assumption? What consequences will this have for the analysis of the data?

Generate a model for the first study.

Is this model also applicable for the second study?

If not, why and what to do?

The last sample has not been measured. Estimate the RAI for this combination.

TIPS:

Modellutvikling:

1. Unngå I størst mulig grad prøver eller variable med missing data – lag så modell
2. Sjekk for outliers: *Score-plott*, *Influence*, *P vs M* (inkl. *RMSEC*) – evt. Ny modell
3. Sjekk antall PC: *Residual VALIDATION variance*
4. Vurder kolinearitett mellom X-er: *Loading plot*, *reg.koef* for flere PC'er sammen
5. Vurder tot. forklart varians for X og Y (angis i *Loading plot*) + *RMSEP* (PvsM)
(evt. *Loading-plot*/"Forklart var. for hver variable" fra en PCA av X-var)
6. Finn signifikante *reg.koef* (evt. for flere PC'er sammen).
Fjern evt. Ikke-signifikante variable en om gangen
ReFit og start på 2 igjen
7. Studer *Respons-overflate* for å forstå sammenhengen mellom faktorer og respons.