

## Data-tolkning / Modellering

### Oppgave 1.

- a. Hva er PLS og hva brukes det til?
- b. Hvilke styrker har denne metoden sammenlignet med MLR?
- c. Beskriv (på et overordnet nivå) hvordan metoden «virker».

### Oppgave 2.

- a. Beskriv 3 metoder for tolkning av data.
- b. Hva er de to vesentligste plottene i PCA og hva forteller de. Gi eksempler på hva de brukes til?
- c. Når kan du bruke MLR og når må du benytte PLS?
- d. Hvilken matematisk metode benyttes for å finne den optimale regresjonslinjen mellom X og Y. Forklar hvordan metoden virker og tegn det gjerne i et X mot Y plott.

### Oppgave 3.

- a. Beskrive kortfattet og i punkter (ikke flere enn 5 punkter) hvordan du går fram ved utvikling av en PLS modell.
- b. Hvilket plott bør benyttes for å studere den kvantitative effekten av to variable og hvorfor nettopp dette plottet.
- c. Hvorfor bør et forsøk som bryter trenden fjernes før endelig modell er bestemt, og i hvilke tilfeller bør en slik prøve beholdes?

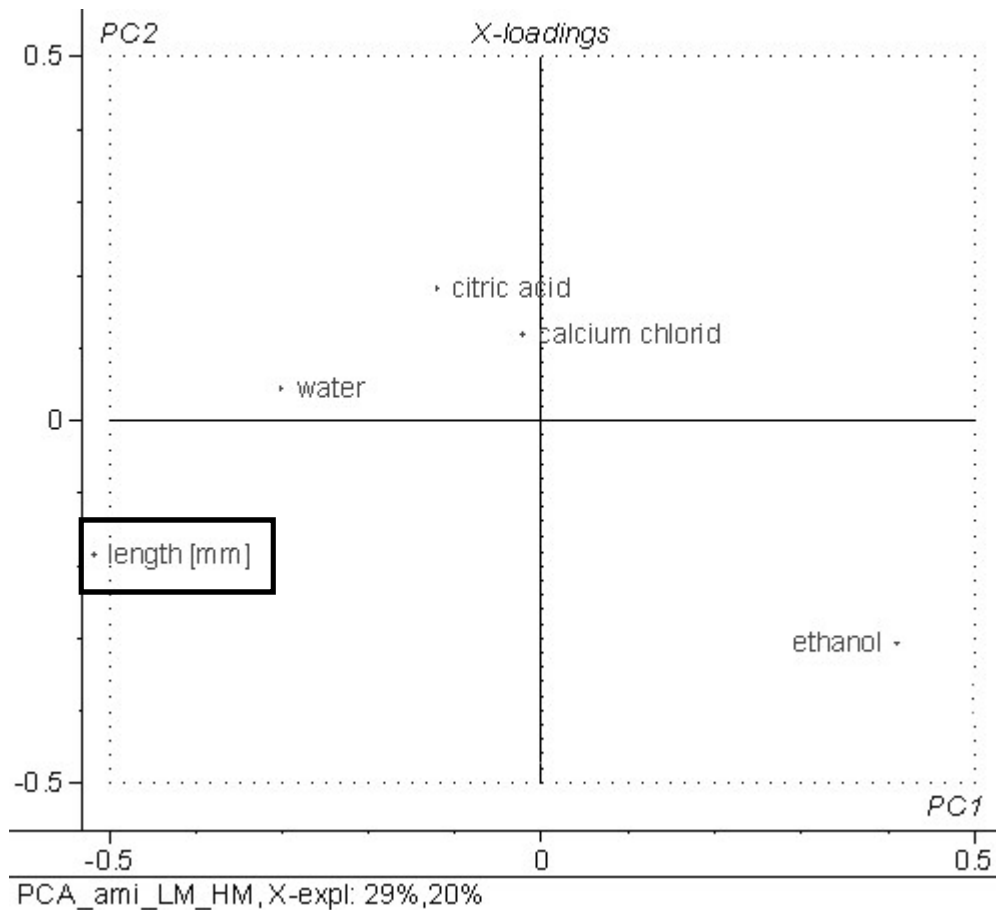
### Oppgave 4.

- a) Hvilke fordeler og ulemper har MLR og PLS
- b) Hvilke plott vil dere studere for avgjøre om en PLS-modell er god, og hvorfor.
- c) Beskriv hvilke funksjoner kryssvalidering har i PLS
- d) Når er kryssvalidering feil metode å benytte og hvorfor

## Div. plott

### Oppgave 5.

I en formuleringsstudie ble det undersøkt hvordan ulike granuleringsvæsker påvirket lengden til de pellets som ble produsert. Det ble granulert med ren etanol, rent vann og vandige løsninger av sitronsyre og kalsiumklorid. I figuren er det vist et plott av loadings fra de to første prinsipalkomponentene etter en PCA av målingene:

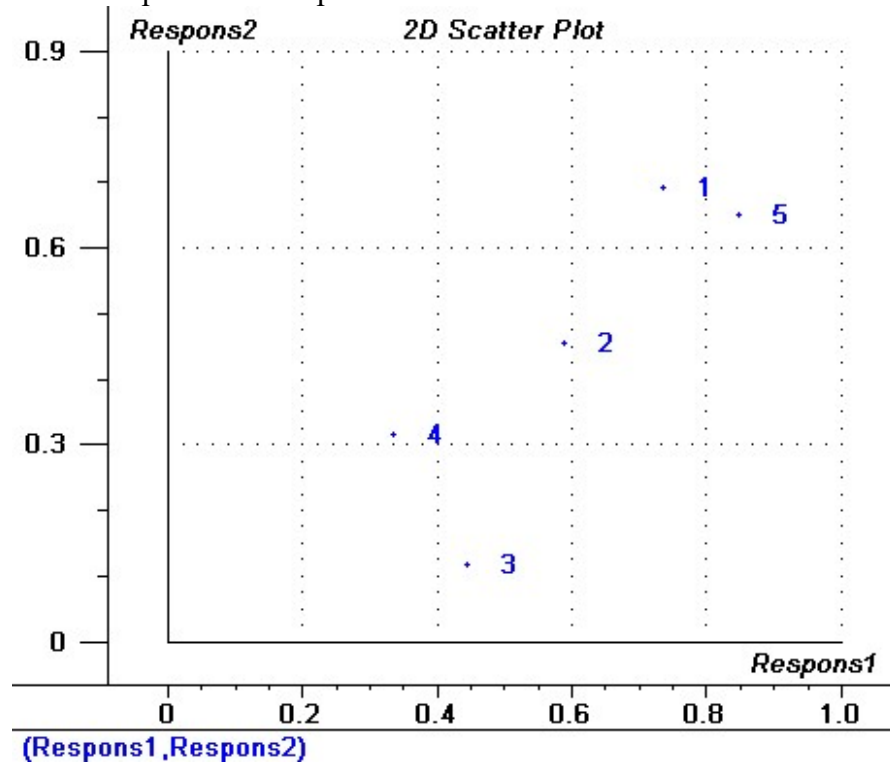


Hvilke av granuleringsvæskene gir lengst og kortest pellets (høyest og lavest verdi for 'length [mm]'?)

Forklar hvordan du kom fram til svarene.

### Oppgave 6.

Et scatter-plott av to responser ser slik ut:



- Tegn et (omtrentlig) score- og loading-plott for en PCA av disse dataene.
- Gi (kort) beskrivelse av hvordan du har gått frem for å generere plottene

## Validering

### Oppgave 7.

- a. Hvorfor er validering essensielt for å finne korrekt modell?
- b. Beskriv 2 aksepterte metoder for å validere en regresjonsmodell samt hvilke fordeler og ulemper disse 2 metodene har.
- c. Det finnes en tredje valideringsmetode som ikke er like god. Hva heter den og hvorfor er den dårligere?
- d. Hvordan finner man signifikante variable i PLS modellering?
- e. Hvilke(t) mål har man for usikkerheten i en PLS modell? Hvordan bruker man dette i evaluering av de predikerte verdiene i responsflaten?

### Oppgave 8.

- a) Hva forteller RMSEC?
- b) Nevn 3 hovedårsaker til hva som bestemmer hvor stor RMSEC blir ?
- c) Hva er forskjellen på RMSEC og RMSEP?

## Avansert modellering

### Oppgave 9.

Tre prosessvariable har blitt undersøkt i storskala produksjon ved hjelp av et 2 nivås fullfaktorielt design (inkludert midtpunkter), med utbytte fra produksjonen som responsvariabel. Det viste seg at to nye prosessvariabler måtte inkluderes i matrisen og at alle 5 variablene og 2 samspill hadde effekt på det målte utbytte. Det var ikke mulig å gjøre flere designede forsøk.

- a. Hvilken regresjonsmetode bør velges for å utvikle en prediksjonsmodell for forventet framtidig utbytte. Kan en slik modell valideres, og eventuelt hvordan.
- b. Hvilke muligheter finnes for å få et omtrentlig estimat av forventet variasjon i framtidige produksjoner.
- c. 5 nye produksjoner viste 2 produksjoner utenfor estimert usikkerhetsområde uten at feil ved produksjonen ble funnet. Hva kan være mulige årsaker?