ISTx1001: Industriell statistikk

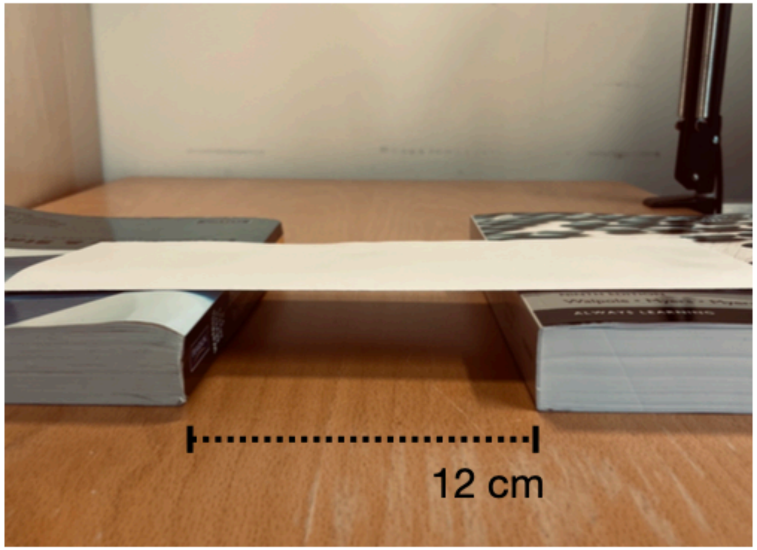
31. oktober 2023

Vi starter klokka 14:15

Timen blir tatt opp :)

Thea Bjørnland, Institutt for matematiske fag, NTNU

Papirbro - hvor mye vekt tåler den?



A: papirtype — grått / hvitt

$x_1 = -1$ $x_1 = +1$

B: avstand — 8 cm / 12 cm

$x_2 = -1$ $x_2 = +1$

C: bretting — 1 brett / 2 brett

$x_3 = -1$ $x_3 = +1$

4.35 gram

2^3 forsøk

2: 2 nivåer per faktor

3: faktorer

$2^3 = 8$ 8 faktorkombinasjoner

Multippel lineær regresjon

Y : respons   kontinuerlig stokastisk variabel

tilnærmet

Samme standardavvik for alle faktorkombinasjoner

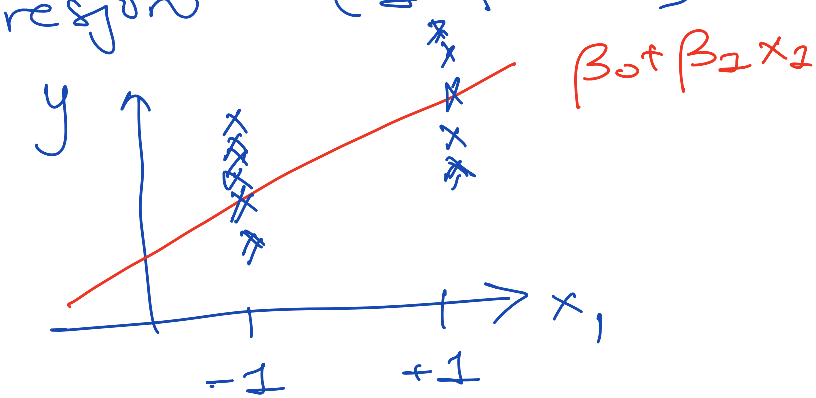
$Y \sim N(\mu, \sigma)$

denne avhenger av faktornivåene

Rep (uke 9)

Enkel lineær regresjon (1 faktor)

$\mu = \beta_0 + \beta_1 x_1$



β_0, β_1 og σ er ukjente, estimere fra data

Multippel lineær regresjon

Flere faktorer (eks 3)

$Y \sim N(\mu, \sigma)$

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 +$$

$$\beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \cancel{\boxed{\beta_{123} x_1 x_2 x_3}}$$

Kan også skrive:

$\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots$

$$Y = \mu + \varepsilon$$

$\varepsilon \sim N(0, \sigma)$

Python: regresjon

datasettet mitt med 16 observasjoner

```
modell = smf.ols('y~x1+x2+x3 + x1*x2 + x2*x3 + x1*x3', data=df).fit()
print(modell.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.947
Model:                            OLS   Adj. R-squared:                  0.911
Method:                 Least Squares   F-statistic:                     26.72
Date:                Thu, 26 Oct 2023   Prob (F-statistic):           3.01e-05
Time:                        16:48:57   Log-Likelihood:                -35.260
No. Observations:                  16   AIC:                             84.52
Df Residuals:                       9   BIC:                             89.93
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     12.2344      0.731     16.745      0.000      10.582      13.887
x1             6.7969      0.731      9.303      0.000       5.144       8.450
x2            -4.0781      0.731     -5.582      0.000      -5.731      -2.425
x3             3.5344      0.731      4.837      0.001       1.882       5.187
x1:x2         -1.9031      0.731     -2.605      0.029      -3.556      -0.250
x2:x3         -0.8156      0.731     -1.116      0.293      -2.468       0.837
x1:x3          2.4469      0.731      3.349      0.009       0.794       4.100
==============================================================================
Omnibus:                        1.224   Durbin-Watson:                   1.642
Prob(Omnibus):                  0.542   Jarque-Bera (JB):                0.454
Skew:                          -0.412   Prob(JB):                        0.797
Kurtosis:                       3.037   Cond. No.                         1.00
==============================================================================
```

Python: regresjon -> samspillseffekter

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	12.2344	0.731	16.745	0.000	10.582	13.887
x1	6.7969	0.731	9.303	0.000	5.144	8.450
x2	-4.0781	0.731	-5.582	0.000	-5.731	-2.425
x3	3.5344	0.731	4.837	0.001	1.882	5.187
x1:x2	-1.9031	0.731	-2.605	$\boxed{0.029}$	-3.556	-0.250
x2:x3	-0.8156	0.731	-1.116	0.293	-2.468	0.837
x1:x3	2.4469	0.731	3.349	0.009	0.794	4.100

β_0 β_1 β_2 β_3

Samspill / interaksjon mellom faktorer

$\hat{\beta}_{12} = -1.9031$ $\qquad \hat{AB} = 2\hat{\beta}_{12} = -3{,}8$ $\qquad p = 0.029 < 0.05$

$\hat{\beta}_{13} = 2.4469$ $\qquad \boxed{\hat{AC} = 2\hat{\beta}_{13} = 4{,}9}$ $\qquad p = 0.009 < 0.05$

hva betyr dette tallet?

Hva betyr en samspillseffekt?

Husk: estimert hovedeffekt av bretting

$\hat{C} = 7{,}1$

i gjennomsnitt 7,1 gram bedre bæreevne ved å endre fra en til to brettinger

→ Er bretting (faktor C) viktigere når jeg bruker hvitt ark eller når jeg bruker grå ark? Eller like viktig for begge typer ark?

Grå ark: $x_1 = -1$

Estimert effekt av å endre fra en til to brettinger når vi bruker grå ark =

$\hat{C} + (-1)\hat{AC} = 7{,}0688 - 4{,}8938 = 2{,}175 \approx 2{,}2g$

Hva betyr en samspillseffekt?

Hvitt ark $x_1 = +1$

Estimert effekt av å endre fra en til to brettinger når vi har hvitt ark

$= \hat{C} + (+1)\hat{AC}$

$= 7{,}0688 + \underbrace{4{,}8938} = 11{,}9262 \approx \underline{12 \text{ gram}}$

dersom $\hat{\beta}_{13}$ hadde vært ≈ 0, ville vi ikke sett disse store forskjellene

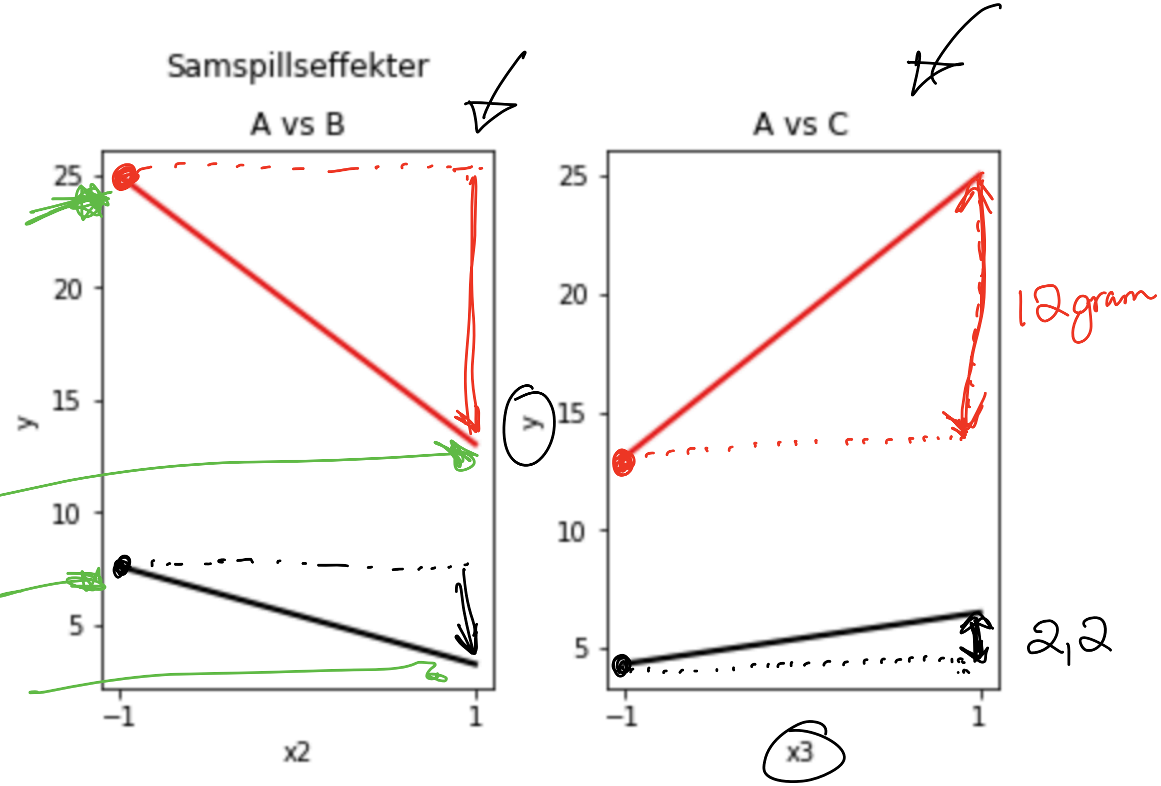
Samspillsplott

Rod : hvitt ark

x1 = +1

Svart : grått ark

x1 = -1

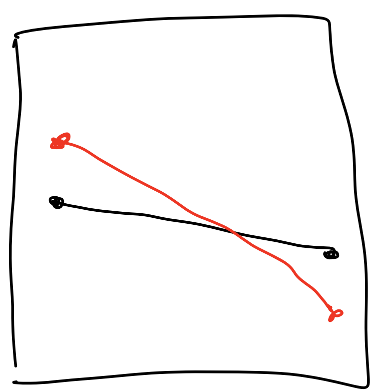


Merk:

Parallelle linjer i samspillsplott ⇒ ingen samspillseffekt

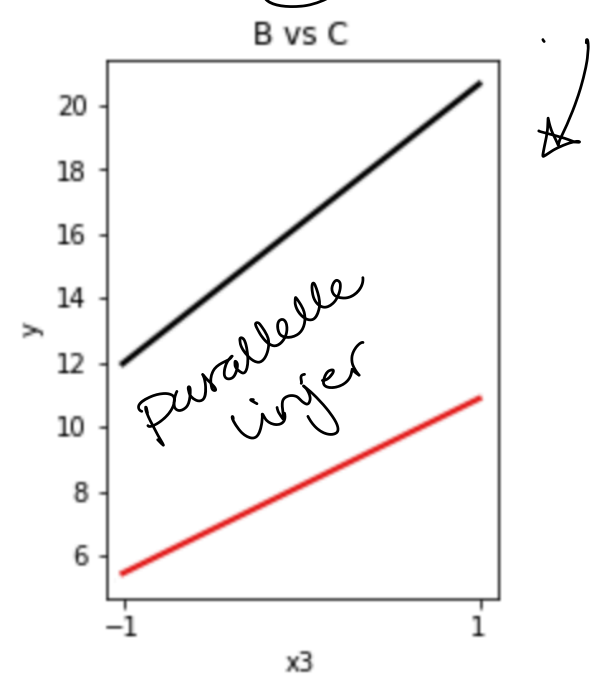
Ikke parallelle linjer

→ samspill



x2 = +1

x2 = -1



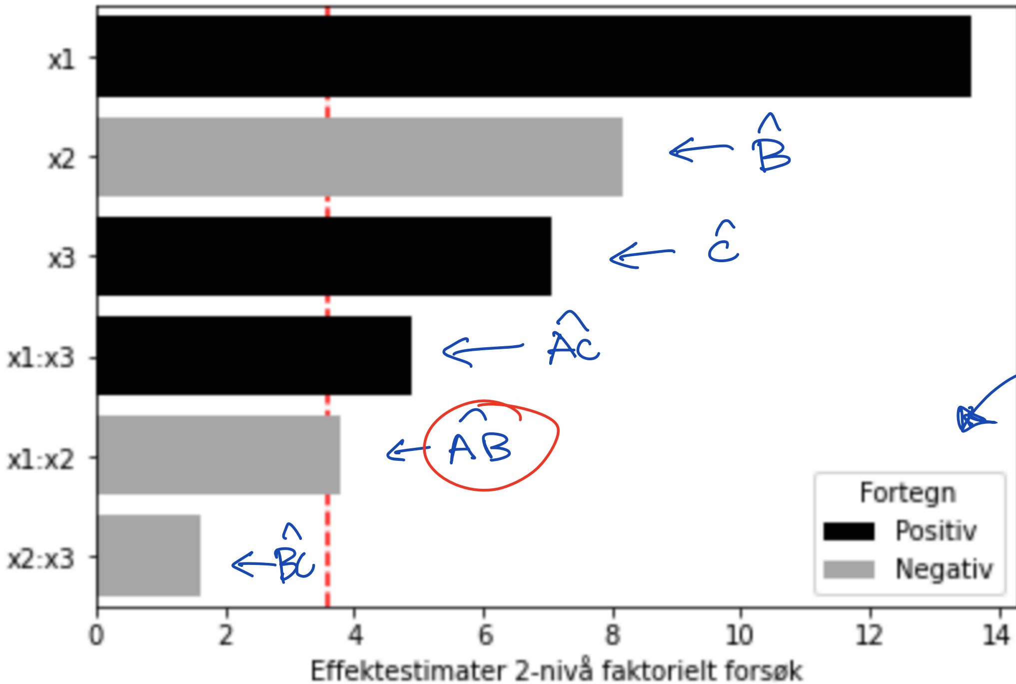
PAUSE TIL

15:15

Paretoplott

kritisk verdi for statistisk signifikans ved signifikansnivå $\alpha = 0.05$

absoluttverdi til estimerte hovedeffekter og samspillseffekter



$\leftarrow \hat{A}$

$\leftarrow \hat{B}$

$\leftarrow \hat{C}$

$\leftarrow \widehat{AC}$

$\leftarrow \widehat{AB}$

$\leftarrow \widehat{BC}$

Kontroll av modellantagelser

* uavhengige forsøk
* normalfordelt respons ← t-testen som gir oss p-verdier (eller den røde linja i Paretoplottet) er definert basert på denne antagelsen

$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$

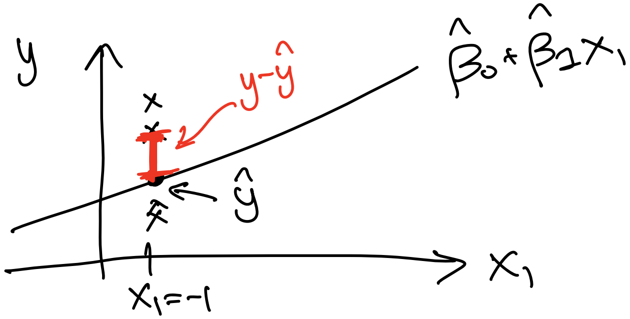
$$Y = \mu + \varepsilon$$

stokastisk

for alle faktorkomb.

$\varepsilon \sim N(0, \sigma)$

$$\hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \hat\beta_3 x_3 + \hat\beta_{12} x_1 x_2 + \dots$$



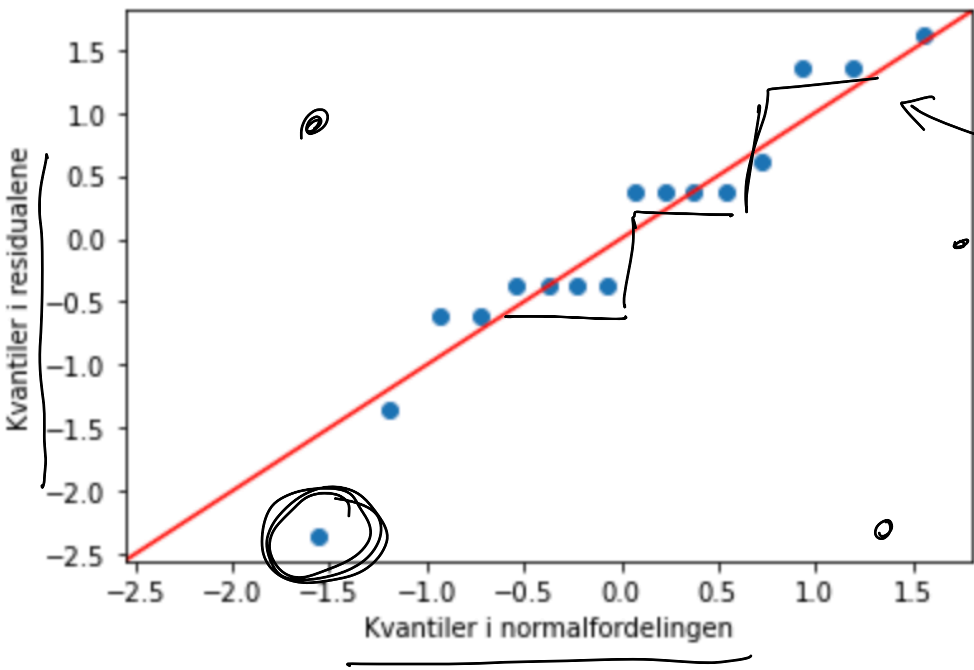
$$\hat e_i = \hat y_i - \hat y_i$$

observasjoner av stokastisk variabel $\underline{\varepsilon \sim N(0, \sigma)}$

Kontroll av modellantagelser

Kontroll av modellantagelser: QQ-plott

Perfekt normalfordelte data: alle blå prikke ligge på den røde linja



"trapp": respons er diskret

Anbefaler dere å gjennomføre steg 1-5 denne uka!

Oppgavebeskrivelse

Målet med prosjektet er å planlegge, gjennomføre og diskutere et 2-nivå faktorielt forsøk. Dere står selv fritt til å velge problemstilling og eksperiment. Det kan være et laboratorieeksperiment eller et problem fra dagliglivet. Her er stegene gruppa skal gjennomføre

1. Bestem dere for en problemstilling som dere kan studere med et 2^k faktorielt forsøk. Hva er responsvariabelen og hva er mulige forklaringsvariabler? Hva er egnede nivåer for forklaringsvariablene? Merk at responsvariabelen Y må være en kontinuerlig stokastisk variabel.
2. Design et forsøk med 16, 24 eller 32 målinger. Bruk enten 3 eller 4 faktorer. Dersom dere bruker 3 faktorer, må dere ha minst ett gjentak.
3. Planlegg innsamling av data.
4. Samle inn data.
5. Analyser data.
6. Skriv en rapport.

Husk: mange eksempler i Hub!

(Uke 9)

Lykke til :)