

ISTx1001: Industriell statistikk

29. oktober 2024

Vi starter klokka 14:15

Timen blir tatt opp

Thea Bjørnland, Institutt for matematiske fag, NTNU

I dag: Fortsette med steg 5 i Del 1 av prosjektet

+ Python

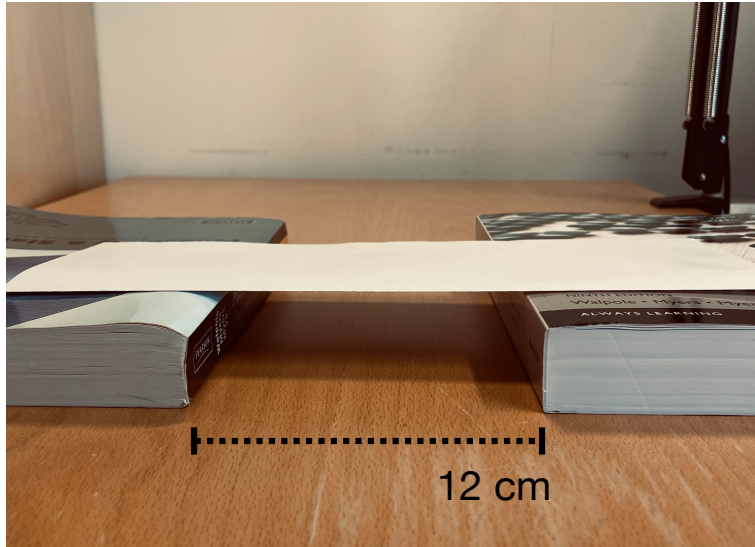
Oppgavebeskrivelse

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.

Anbefaler at dere har samlet inn data og forsøkt å analysere dem før veiledning torsdag. Bruk Jupyter-notatbøker med eksempler i huben

Papirbro - hvor mye vekt tåler den?



4.35 gram

Responsvariabel: broas bæreevne [gram]

Faktor A : Type ark.

Lavt nivå: Grått ark ($x_1 = -1$)

Høyt nivå: Hvitt ark ($x_1 = +1$)

Faktor C : Bretting

Lavt nivå: Brettet en gang ($x_3 = -1$)

Høyt nivå: Brettet to ganger ($x_3 = +1$)

Faktor B : Avstand

Lavt nivå: 8 cm ($x_2 = -1$)

Høyt nivå: 12 cm ($x_2 = +1$)

Resultater



$x_1 (A)$	$x_2 (B)$	$x_3 (C)$
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

y1	y2
8.70	4.35
17.40	17.40
0.00	4.35
8.70	8.70
8.70	8.70
30.45	34.80
4.35	4.35
13.05	21.75

Resultater

Bretting

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

y1	y2
8.70	4.35
17.40	17.40
0.00	4.35
8.70	8.70
8.70	8.70
30.45	34.80
4.35	4.35
13.05	21.75

Snitt bæreevne en brett = 8.7 g

Snitt bæreevne to brett = 15.76875 g

Hovedeffekt bretting = 15.76875 - 8.7 = 7.06875 g

↑
høyt
nivå

↑
lavt
nivå

hovedeffekt

Resultater

avstand
lavt nivå (8cm)

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

v1	v2
8.70	4.35
17.40	17.40
0.00	4.35
8.70	8.70
8.70	8.70
30.45	34.80
4.35	4.35
13.05	21.75

Snitt bæreevne en brett = 11.9625 g

Snitt bæreevne to brett = 20.6625 g

Effekt av brekking når avstand er på lavt nivå = $20.6625 - 11.9625 = 8.7 \text{ g}$

Resultater

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$	y1	y2	
-1	-1	-1	8.70	4.35	Snitt bæreevne en brett = 11.9625 g
+1	-1	-1	17.40	17.40	
-1	+1	-1	0.00	4.35	Snitt bæreevne en brett = 5.4375 g
+1	+1	-1	8.70	8.70	
-1	-1	+1	8.70	8.70	Snitt bæreevne to brett = 20.6625 g
+1	-1	+1	30.45	34.80	
-1	+1	+1	4.35	4.35	Snitt bæreevne to brett = 10.8730 g
+1	+1	+1	13.05	21.75	

Effekt av brekking når avstand er på lavt nivå = $20.6625 - 11.9625 = 8.7 \text{ g}$

Effekt av brekking når avstand er på høyt nivå = $10.8730 - 5.4375 = 5.4375 \text{ g}$

Avstand
12 cm

Resultater

$x_1(A)$	$x_2(B)$	$x_3(C)$	y1	y2	
-1	-1	-1	8.70	4.35	Snitt bæreevne en brett = 11.9625 g
+1	-1	-1	17.40	17.40	
-1	+1	-1	0.00	4.35	Snitt bæreevne en brett = 5.4375 g
+1	+1	-1	8.70	8.70	
-1	-1	+1	8.70	8.70	Snitt bæreevne to brett = 20.6625 g
+1	-1	+1	30.45	34.80	
-1	+1	+1	4.35	4.35	Snitt bæreevne to brett = 10.8730 g
+1	+1	+1	13.05	21.75	

Effekt av breting når avstand er på lavt nivå = $20.6625 - 11.9625 = 8.7 \text{ g}$

Effekt av breting når avstand er på høyt nivå = $10.8730 - 5.4375 = 5.4375 \text{ g}$

Samspillseffekt breting og avstand = $0.5 * (5.4375 \text{ g} - 8.7) = -1.63125 \text{ g}$

Resultater

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

y1	y2
8.70	4.35
17.40	17.40
0.00	4.35
8.70	8.70
8.70	8.70
30.45	34.80
4.35	4.35
13.05	21.75

Hovedeffekt bretteing = 7.06875 g

Samspillseffekt bretteing og avstand = -1.63125 g

* Modifisering av brette-effekt når broa er kort (lavt nivå): 7.06875 + (-1)* -1.63125 = 8.7 g

hovedeffekt

samspillseffekt

* Modifisering av brette-effekt når broa er lang (høyt nivå): 7.06875 + (+1)* -1.63125 = 5.4375 g

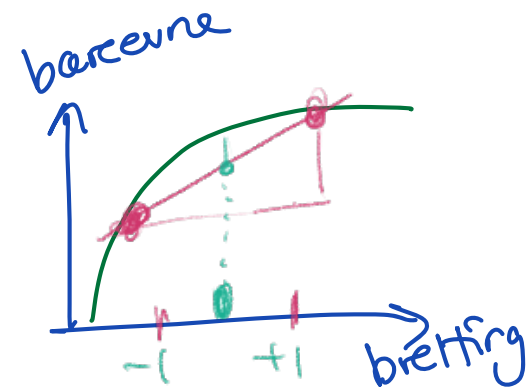
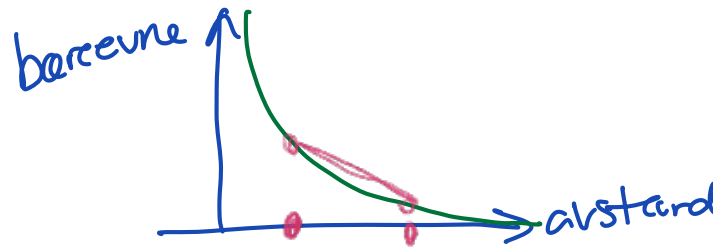
$x_2 = +1$

2-nivå faktorielle forsøk: hvorfor er det så lurt med dette oppsettet?

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

1. x-ene (forklaringsvariablene) har bare to tallverdier (og disse er like for alle)

→ lineære sammenhenger



2. x-ene er ukorrelerte (ortogonale kolonner)

kan studere en faktor av gangen

3. alle kombinasjoner av x-ene er med

kan undersøke om det finnes
samspillseffekter.

Multipel lineær regresjon

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \varepsilon$$

$E(Y)$

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

feil-ledd
stykket

NB: kun opptatt av $E(Y)$ ved
x-verdier +1 eller -1.

(har ikke grunnlag for å studere noe annet)

Python: format på dataene

```
colnames = ["x1", "x2", "x3", "y"]
rep1 = pd.DataFrame(np.column_stack([X, y1]), columns=colnames)
rep2 = pd.DataFrame(np.column_stack([X, y2]), columns=colnames)

df = pd.concat([rep1, rep2])

print(df)
```

	x1	x2	x3	y
0	-1.0	-1.0	-1.0	8.70
1	1.0	-1.0	-1.0	17.40
2	-1.0	1.0	-1.0	0.00
3	1.0	1.0	-1.0	8.70
4	-1.0	-1.0	1.0	8.70
5	1.0	-1.0	1.0	30.45
6	-1.0	1.0	1.0	4.35
7	1.0	1.0	1.0	13.05
0	-1.0	-1.0	-1.0	4.35
1	1.0	-1.0	-1.0	17.40
2	-1.0	1.0	-1.0	4.35
3	1.0	1.0	-1.0	8.70
4	-1.0	-1.0	1.0	8.70
5	1.0	-1.0	1.0	34.80
6	-1.0	1.0	1.0	4.35
7	1.0	1.0	1.0	21.75

Rep 1

Rep 2.

Python: regresjon

← t-test

	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

$$\hat{\beta}_3 = 3.5344 \rightarrow 2\hat{\beta}_3 = 7.068 \dots \text{ hovedeffekt}$$

$$H_0: \beta_3 = 0$$

↑
ingen effekt
av brekking

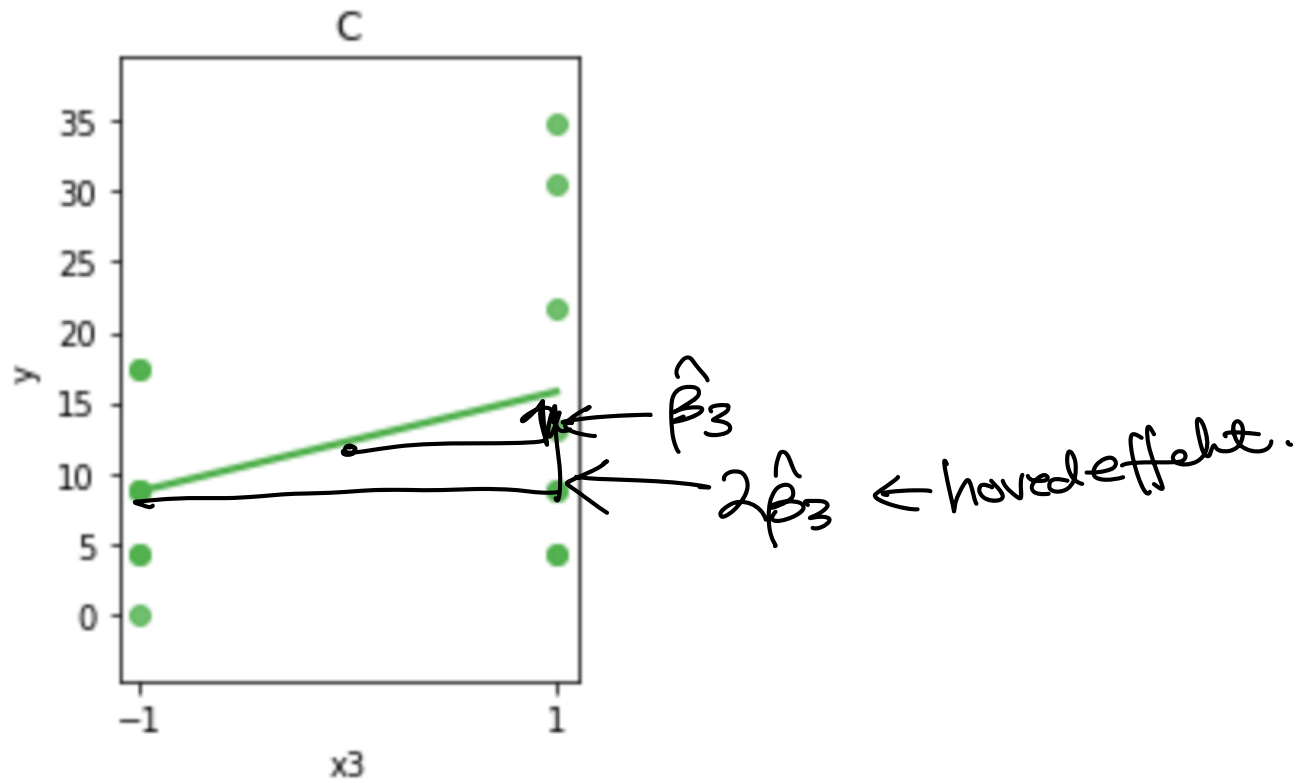
$$H_1: \beta_3 \neq 0$$

↑
effekt
av brekking

$$p\text{-verdi} = 0.001 < \uparrow 0.05$$

→ signifikant
effekt av
brekking.

Hovedeffektsplott



Python: regresjon

	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

Samspill / interaksjon
mellom brekking og avstand

$$\hat{\beta}_{23} = -0.8156$$

$$\hookrightarrow 2\hat{\beta}_{23} = -1.6312$$

avstand på høyt nivå $x_2 = +1$

↳ effekten av brekking:

$$2\hat{\beta}_3 + (+1)2\hat{\beta}_{23}$$

$$H_0: \beta_{23} = 0$$

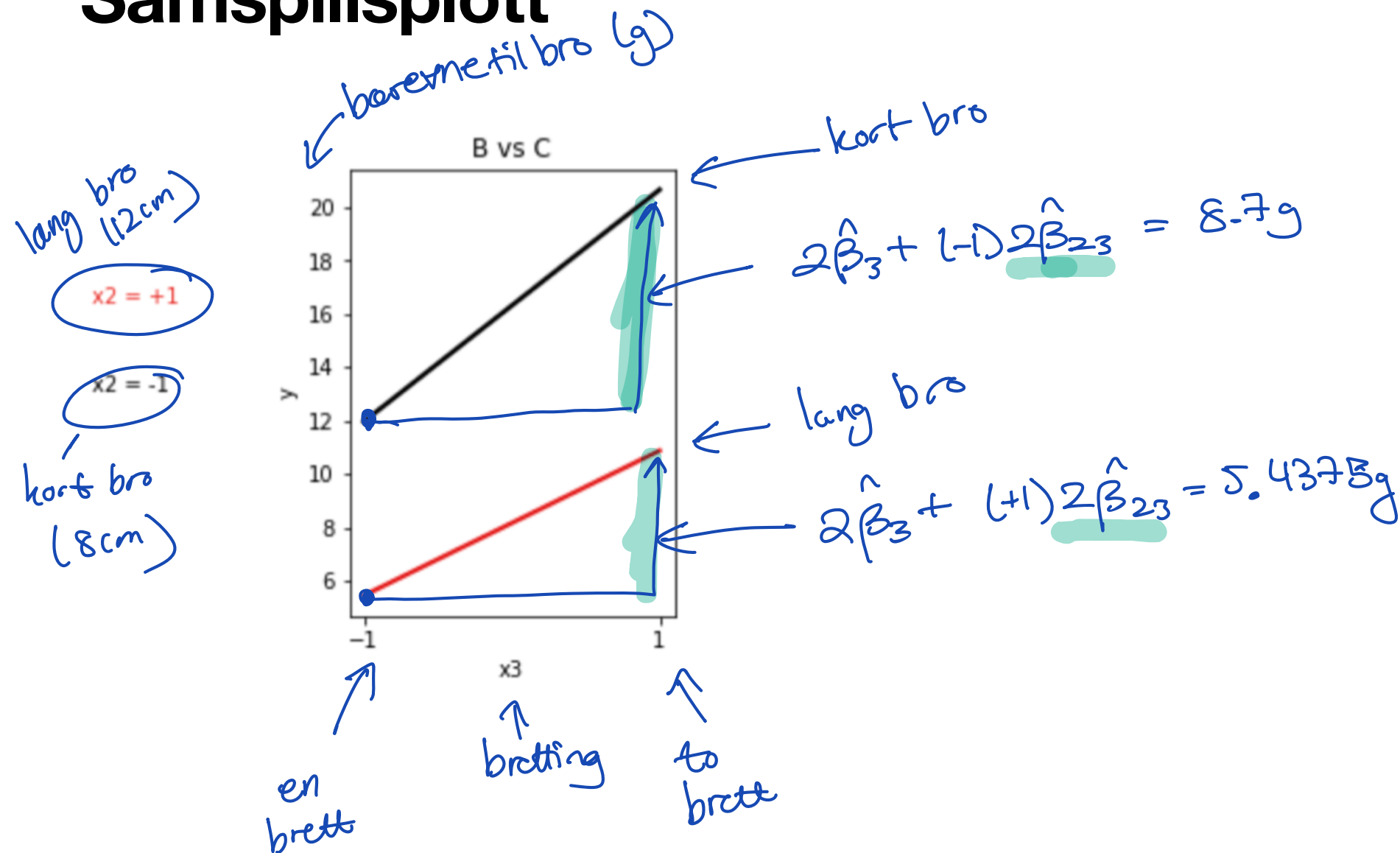
(ingen samspillseffekt)

$$H_1: \beta_{23} \neq 0$$

(samspillseffekt)

p-verdi = 0.293 → kan ikke forkaste H_0

Samspillsplott



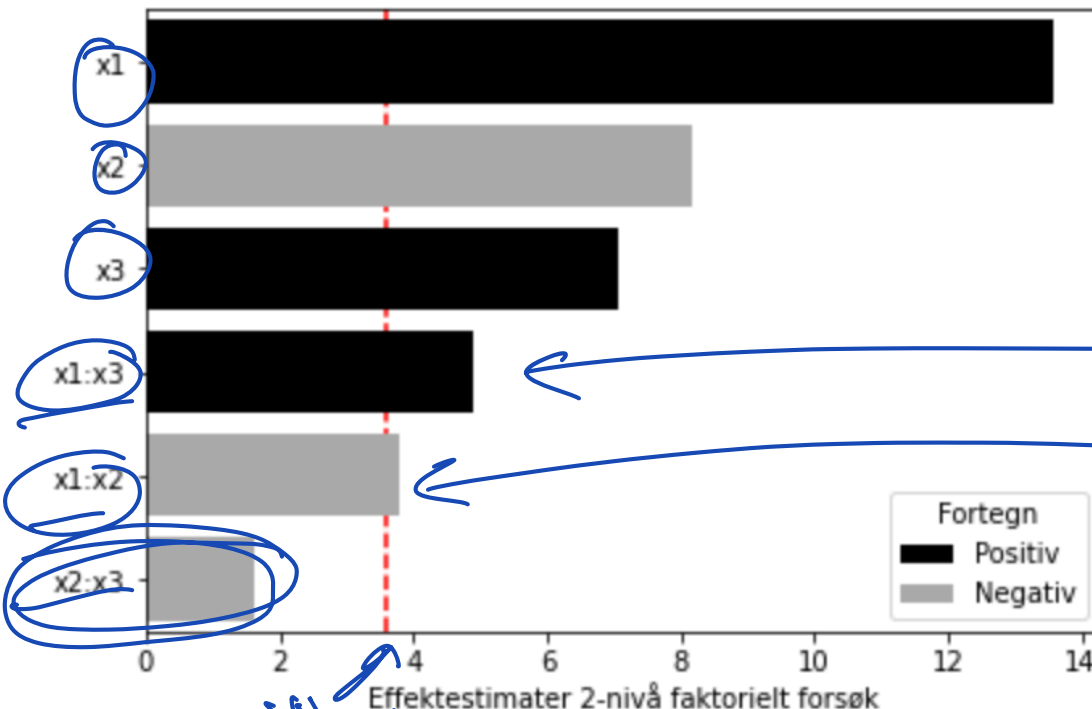
Python: regresjon - alle faktorer

merk: her ikke med $x_1 x_2 x_3$

papirtype
avstand
bretting

	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

Paretoplott



signifikans

$$2\hat{\beta}_1$$

hovedeffekt
av papirtype

$$2\hat{\beta}_2$$

hovedeffekt
avstand.

$$2\hat{\beta}_{13}$$

samspilleffekt
papirtype og bretting.

$$2\hat{\beta}_{12}$$

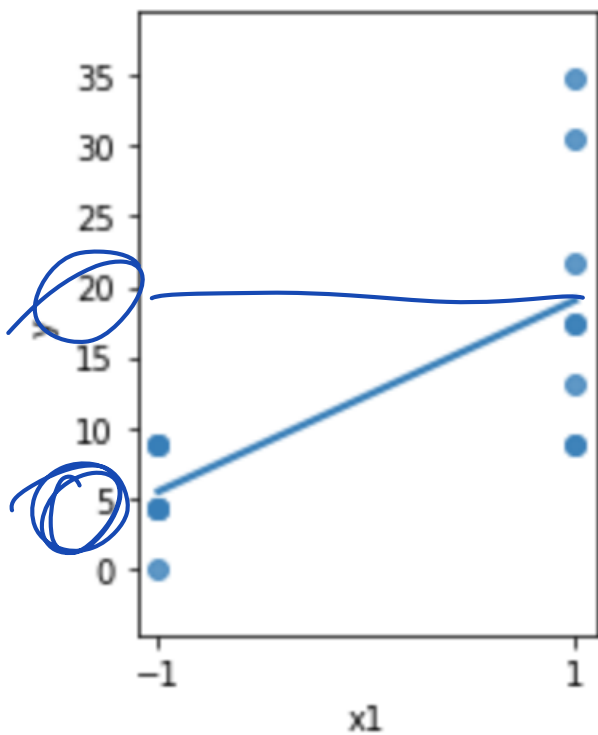
samspill
papirtype og avstand.

Hovedeffektsplott

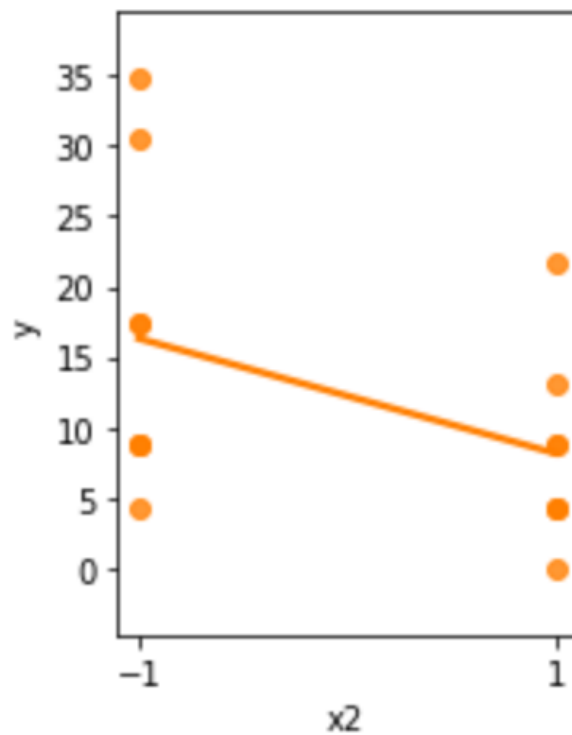
højest
signifikante

Hovedeffekter

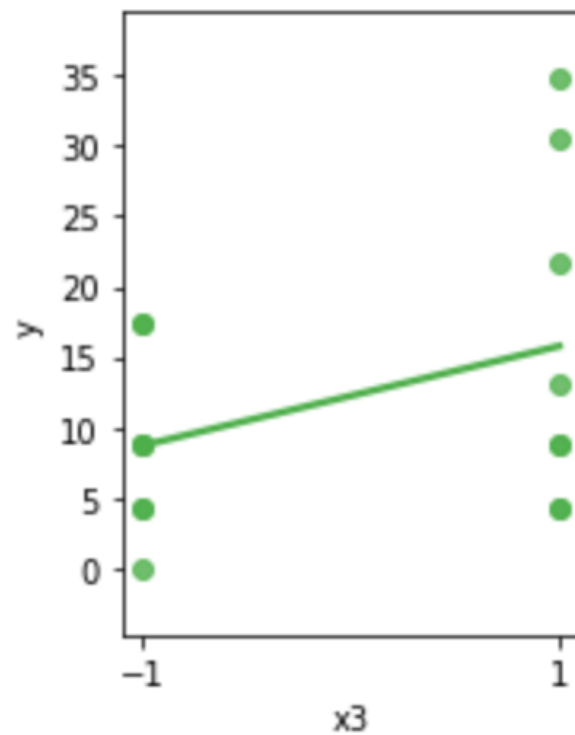
A



B



C



papir

avstand

bredding

PAUSE TIL
15:15.

Samspillsplott

hvitte papir

$x1 = +1$

$x1 = -1$

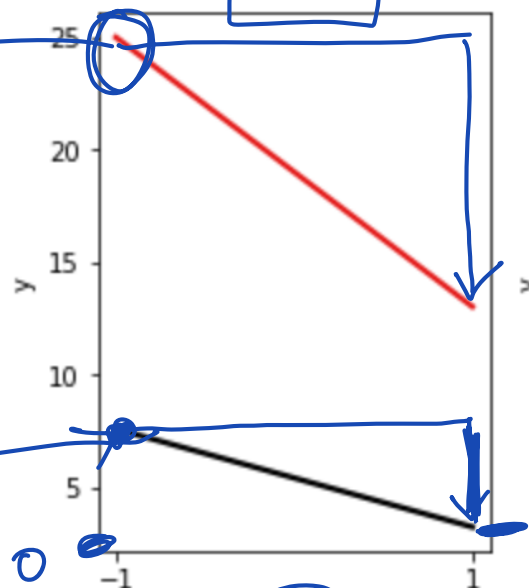
grønt papir

8cm

avstand

Samspillseffekter

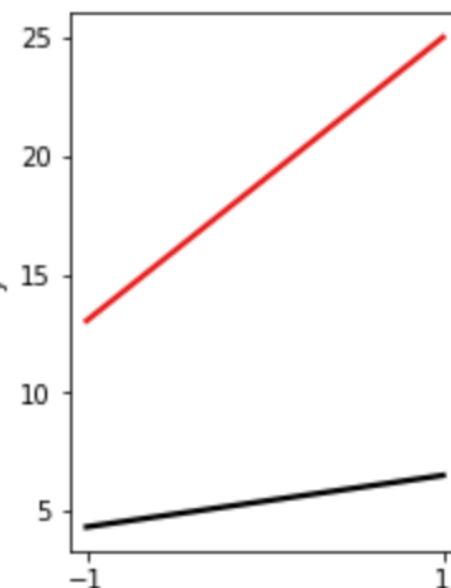
A vs B



$x2 = +1$

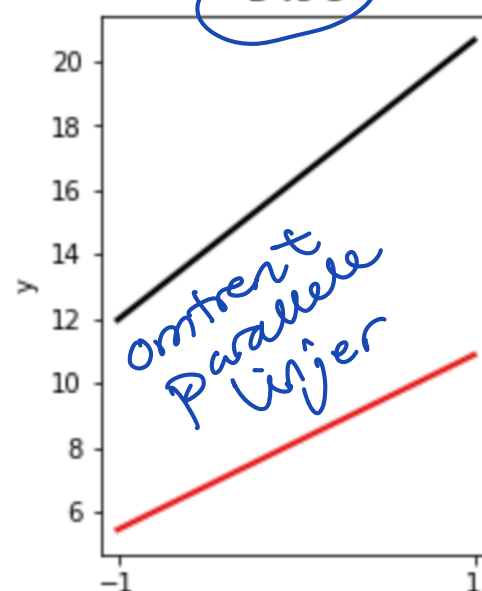
$x2 = -1$

A vs C



$x3$

B vs C



$x3$

Kontroll av modellantagelser

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{23} x_2 x_3 + \beta_{13} x_1 x_3 + \varepsilon$$

$$\varepsilon \sim \underline{N(0, \sigma)}$$

• Normalantagelse på ε

$$i = 1, \dots, n \quad (n=16)$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

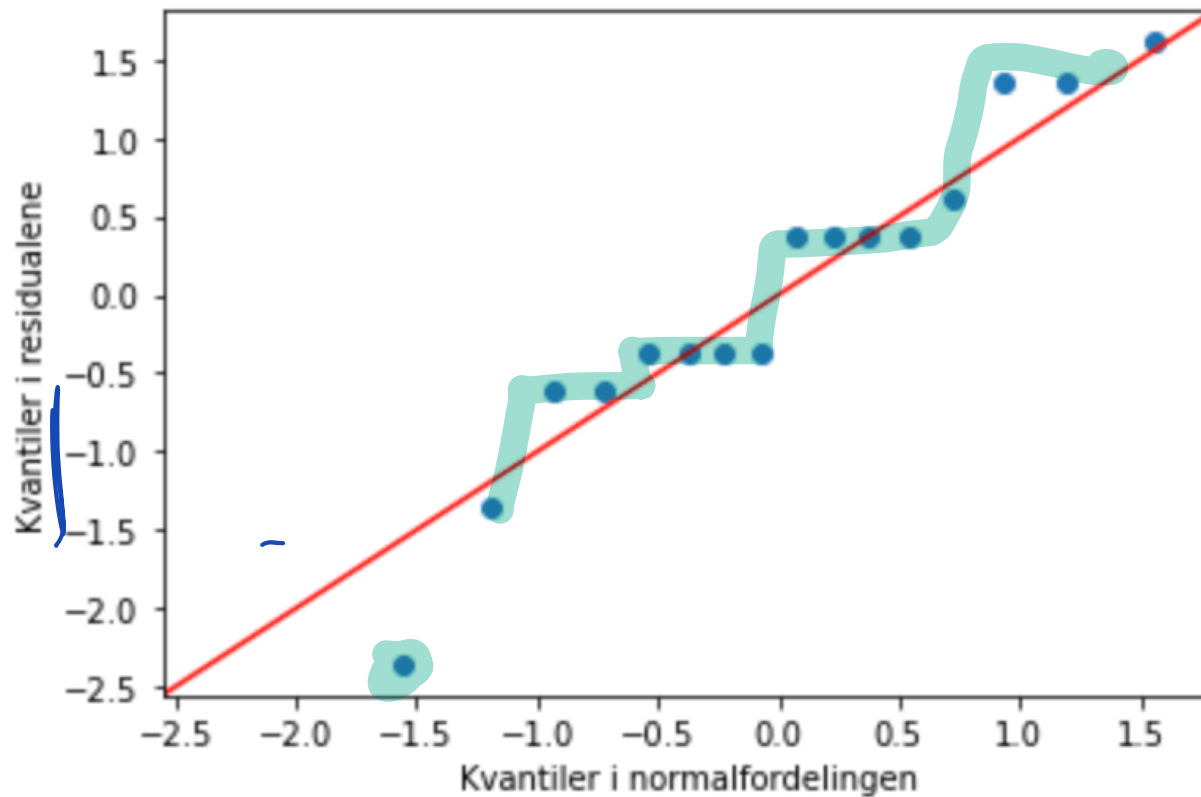
← Skal oppføre seg som
om de kommer fra
Summe Normalfordeling

↑
(summe varians og forventning)

Kontroll av modellantagelser

QQ-plott. (quantile - quantile)

← ligger de blå prikkene på rød linje?



Tips!

Det ligger flere notatbøker på huben der jeg har brukt Python for å estimere effekter i 2-nivå faktorielle forsøk :)

Litt om sensur

En *god* besvarelse innebærer at studentene har svart på alle punktene i prosjektbeskrivelsen. Kandidatene viser forståelse ved f.eks. å forklare resultater og metode med egne ord, og tilpasset deres eget forsøk.

En *middels* besvarelse består stort sett av de riktige tabeller og figurer, men viser ikke tilstrekkelig forståelse for temaene som dekkes i prosjektmodulen 1001.

En *svak* besvarelse inneholder ikke alle punktene fra prosjektbeskrivelsen og viser svært litt forståelse for temaene som dekkes i prosjektmodulen 1001.

Noen tips:

- Les prosjektbeskrivelsen nøye!
- Begrunn valg av faktorer og nivåer dere har brukt i forsøket deres
- Randomiser rekkefølgen
(og forklar hvordan dere gjorde dette i praksis)
- Skriv opp regresjonsmodellen dere bruker.
Har dere med samspill mellom tre eller fire faktorer?
- Inkluder modell-sammendrag (fra Python) og alle figurer
- Forklar hva dere lærer om forsøket fra sammendrag og figurer, gi eksempler. Forstår dere forskjellen på signifikant effekt og praktisk relevant effektstørrelse?
Hvordan tolker dere samspillseffekter?