

ISTx1001: Industriell statistikk

30. oktober 2023

Vi starter klokka 15:15

Timen blir tatt opp :)

Thea Bjørnland, Institutt for matematiske fag, NTNU

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

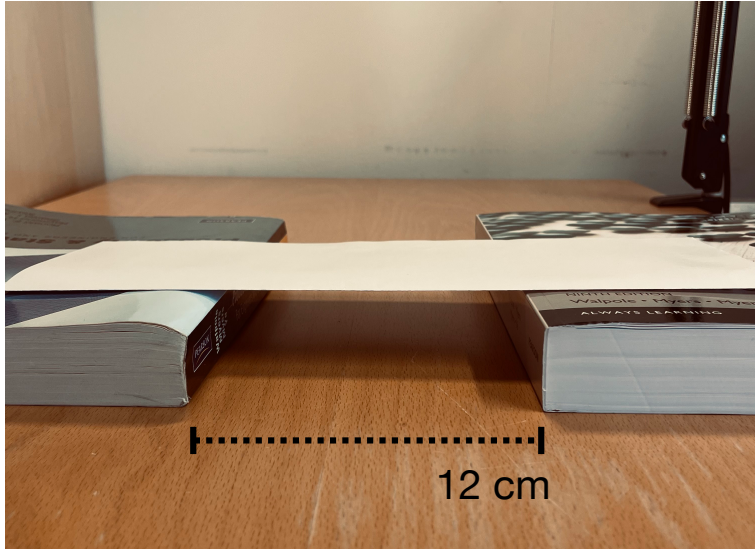
+ Python

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. ← i dag og i morgen
6. Skriv en rapport.

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: bretteing 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 faktor kombinasjoner

Designmatrise

(rep)

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

- alle faktor kombinasjoner gjennomføres
- randomisert rekkefølge

$2^3 = 8$ rader (faktor kombinasjon)

Resultater

målt bæreevne i første repetisjon
↓

andre repetisjon
↓

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

Multipel lineær regresjon

Y : responsvariabel (stokastisk, kontinuerlig)

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Slipningspunkt
↑
 $+ \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3$

$+ \beta_{123} X_1 X_2 X_3$
Ser bort ifra
samspill for
mer enn to
faktorer

Husk: β_1 = forventet endring i Y
når X_1 øker med 1

→ $2\beta_1$ = forventet endring i Y
når X_1 går fra -1 til +1

[faktor A går fra lavt til høyt nivå]

{ bytter papirtype fra grått til hvitt }

Multippel lineær regresjon

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

merk!

Python: regresjon

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

} Oke 9

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 -> hovedeffekter

	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

Hovedeffekt papir type
Hoved effekt avstand
Hoved effekt breiting

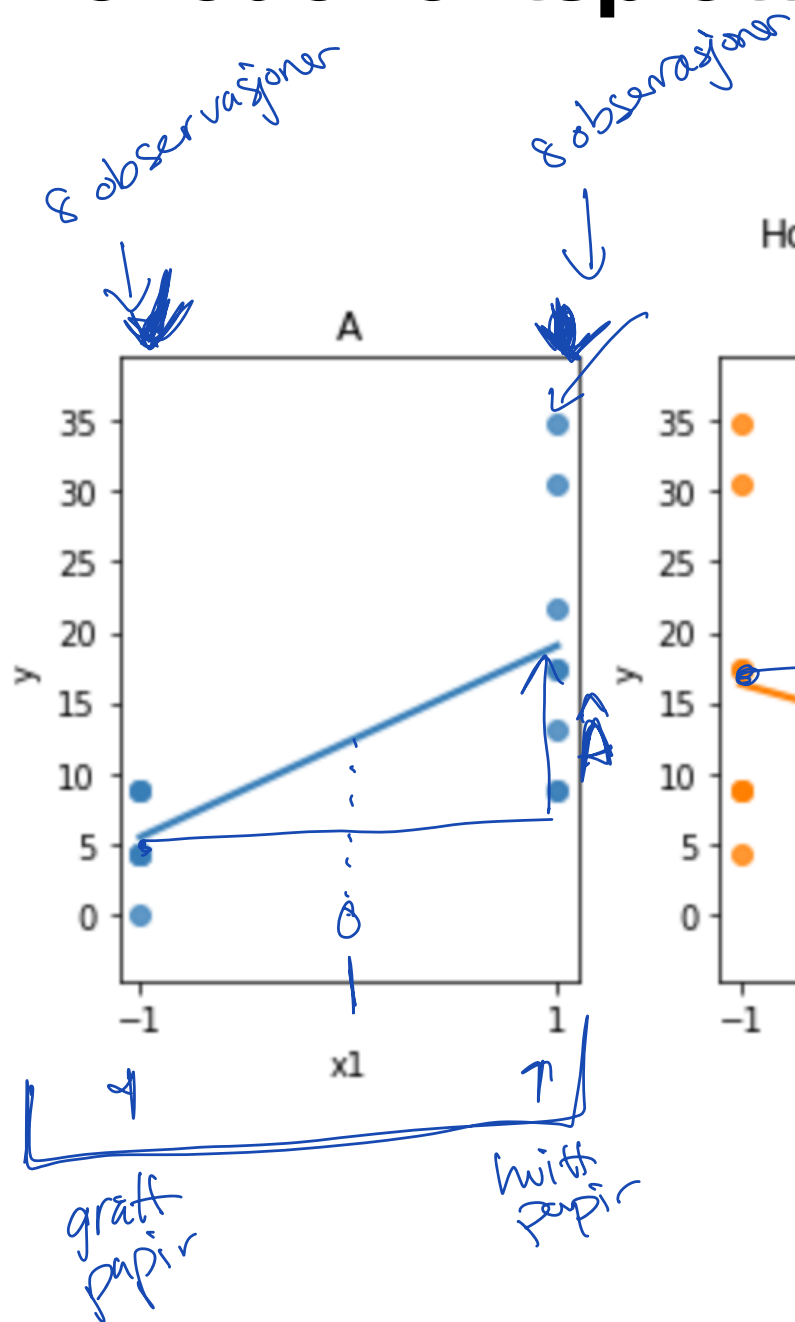
$$\hat{A} = 2\hat{\beta}_1 = 2 \cdot 6,7969 = 13,5938$$

$$\hat{B} = 2\hat{\beta}_2 = 2 \cdot (-4,0781) = -8,1562$$

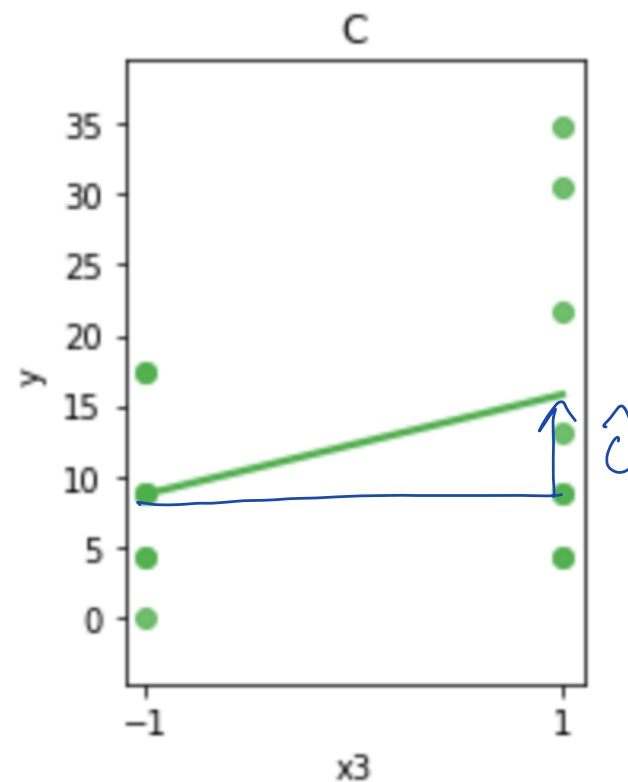
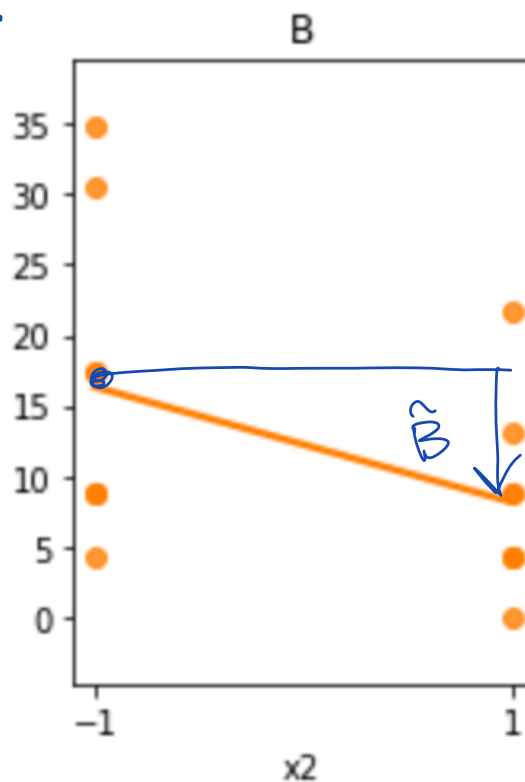
$$\hat{C} = 2\hat{\beta}_3 = 2 \cdot 3,5344 = 7,0688 \quad (?)$$

Hovedeffektsplott

(Se huben for
kode for å lage
slike plott)



Hovedeffekter



dårligere
bæreevne når
vi går fra lavt
til høyt nivå

Python: regresjon -> hovedeffekter

	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

Eks

$$H_0: \beta_1 = 0$$

(ingen effekt
av papirtype)

$$H_1: \beta_1 \neq 0$$

(effekt)

husk: ved signifikansnivå
 $\alpha = 0.05$ så forkaster
vi H_0 til fordel for H_1
dersom p-verdien < 0.05

T-test

Vi kan lese av resultatet fra
t-testen ved å se på p-verdier
i tabellen "P>|t|"

$$p = 0.000?$$
$$p < 0.0005$$

Konkl: Å bytte fra
grått til hvitt papir har
en statistisk signifikant
sammenheng med
forventet bærsevne.

Tips!

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

| morgen 14:15 - 16:00

- Samspillseffekter
- sjekke av modell antagelser