

ISTx1001: Industriell statistikk

28. oktober 2024

Vi starter klokka 14:15

Timen blir tatt opp :)

Thea Bjørnland, Institutt for matematiske fag, NTNU

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

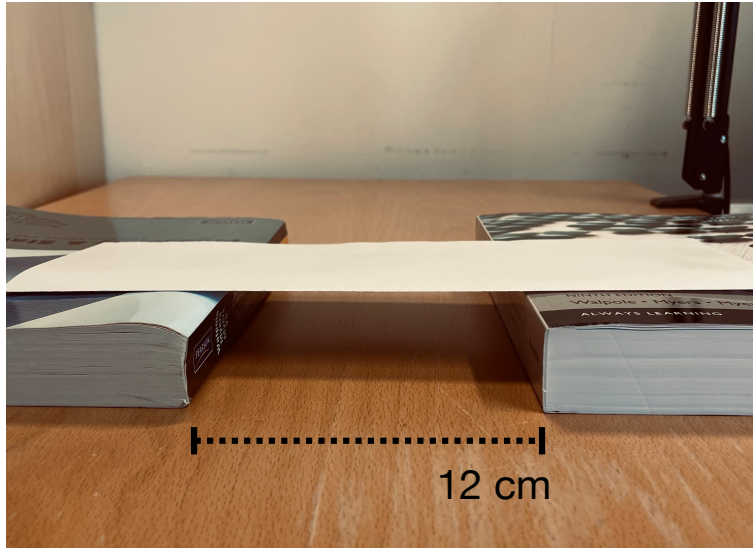
+ Python

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.
6. Skriv en rapport.

Papirbro - hvor mye vekt tåler den?



4.35 gram

A: Papirtype (grått/hvittark)
 $x_1 = -1, x_1 = +1$

B: avstand (8cm, 12cm)
 $x_2 = -1, x_2 = +1$

C: bredding (1 Brett, 2 Brett)
 $x_3 = -1, x_3 = +1$

2³ ← 3 faktorer
↑
- forsøk

2-nivå forsøk,
hver faktor på 2 nivåer

Respons: bæreevne i gram

Designmatrise

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

← grøtt ark,
8cm avstand,
1 brett

← hvitt ark,
12cm avstand,
2 brett

Designmatrise

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

2. x-ene er ukorrelererte (ortogonale kolonner)

3. alle kombinasjoner av x-ene er med

2-nivå
full-faktorielt forsøk

-> vi må randomisere rekkefølgen når vi gjør forsøk

-> vi bør gjøre gjentak av forsøket

→ dette har delvis med frihetsgrader å gjøre.

NB: I morgen snakker vi mer om hvorfor det er så lurt på gjøre forsøk på akkurat denne måten

Resultater

observert
bæreevne (g)
i første 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

viser
at y er stokastisk.

↑
andre repetisjon
(gjentak)

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

Effekt av faktor C: Bretting

- snitt bæreevne for bretting på høyt nivå
2 brett

$$\bar{y} = \frac{8.7 + 30.45 + 4.35 + 13.05 + \dots}{8}$$

$$= 15.76875 \text{ g}$$

- snitt bæreevne for bretting på lavt nivå
1 brett

$$\bar{y} = 8.7 \text{ g}$$

$$\rightarrow \text{forskjell} = 15.76875 \text{ g} - 8.7 \text{ g} = 7.06875 \text{ g}$$

$\approx 7.1 \text{ g}$ økt bæreevne ved å gå fra en til to brett (lavt til høyt nivå)

Resultater

← avstand ← bretteing

$x_1 (A)$	$x_2 (B)$	$x_3 (C)$	y1	y2
-1	-1	-1	8.70	4.35
+1	-1	-1	17.40	17.40
-1	+1	-1	0.00	4.35
+1	+1	-1	8.70	8.70
-1	-1	+1	8.70	8.70
+1	-1	+1	30.45	34.80
-1	+1	+1	4.35	4.35
+1	+1	+1	13.05	21.75

Samspillseffekt bretteing og avstand

① Avstand lavt nivå (8cm) 

• Brett høyt nivå : $\bar{y} = 20.6625g$

• Brett lavt nivå : $\bar{y} = 11.9625g$

⇒ Effekt av bretteing når avstand er på lavt nivå =
 $20.6625g - 11.9625g = 8.7g.$

② Avstand høyt nivå 

• Brett høyt nivå $\bar{y} = 10.873g$

• Brett lavt nivå $\bar{y} = 5.4375g$

Effekt av bretteing når avstand er på høyt nivå =

$$10.873g - 5.4375g = 5.4375g$$

Fra ① og ② ser vi at det kan være bedre effekt av bretteing når det er kort avstand mellom bokene

Samspillseffekt

$$\frac{1}{2}(5.4375g - 8.7g) = -1.63125g$$

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

Hovedeffekt bretting

$\approx 7.7g$

↳ estimert effekt av å endre fra en til to brett.

Men, dersom broa er lang modifieres hovedeffekt:

$$7.068... + (-1.63125) = 5.4375g$$

Dersom broa er kort

$$7.068... + (-1)(-1.63125) = 8.7g$$

hovedeffekt x_3 (bretting)
 $x_2 = -1$
samspilleffekt

Multippel lineær regresjon

husk:

Forventet borseigne (gram)

støy/
variasjon

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{23} X_2 X_3 + \beta_{13} X_1 X_3 + \epsilon$$

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

"hovedeffekter"

"Samspill/
interaksjon"

NB: har ikke tatt
med i modellen
 $X_1 X_2 X_3$

husk: β_1 = forventet endring i
 Y når X_1 øker med 1,
å øke fra -1 til +1 er
en endring på 2,

→ derfor $2\beta_1$ (osv) er 'effekter'

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

repetisjon 1

Repetisjon 2

Python: regresjon

hovedeffekter

Samspill

```
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		

$$16 - 7 = 9$$

t₉

	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

```
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

	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 x3 (bretting)

$$\hat{\beta}_3 = 3.5344 \Rightarrow 2\hat{\beta}_3 = 7.068 \dots$$

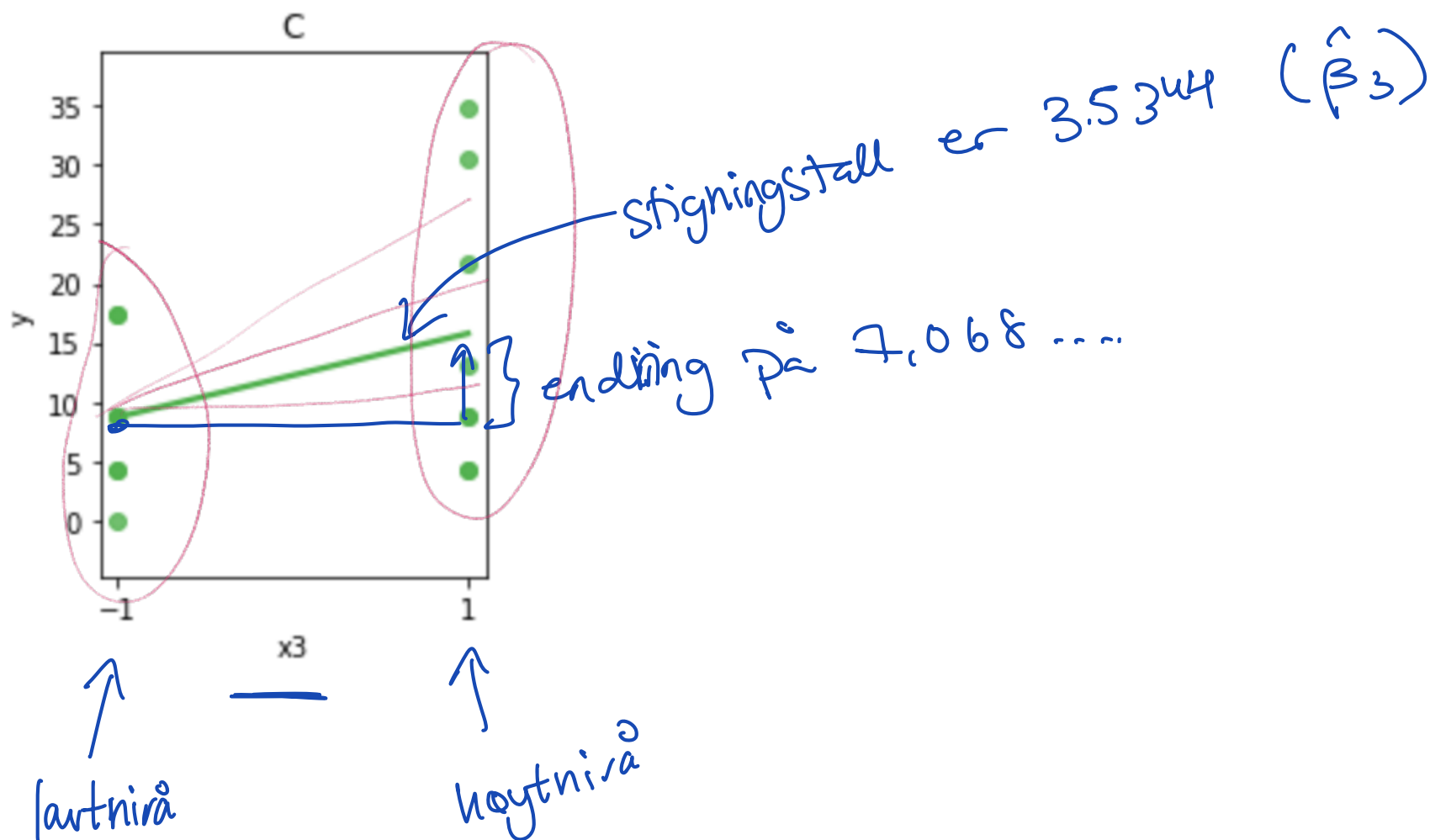
p-verdi fra t-test = 0.001 (< 0.05)

↳ signifikant hovedeffekt av bretting.

← Samme effekt av
bretting som vi
regnet manuelt!

NB!
bare gyldig
dersom
normalantakelse
på ϵ (støy)
stemmer

Hovedeffektsplott



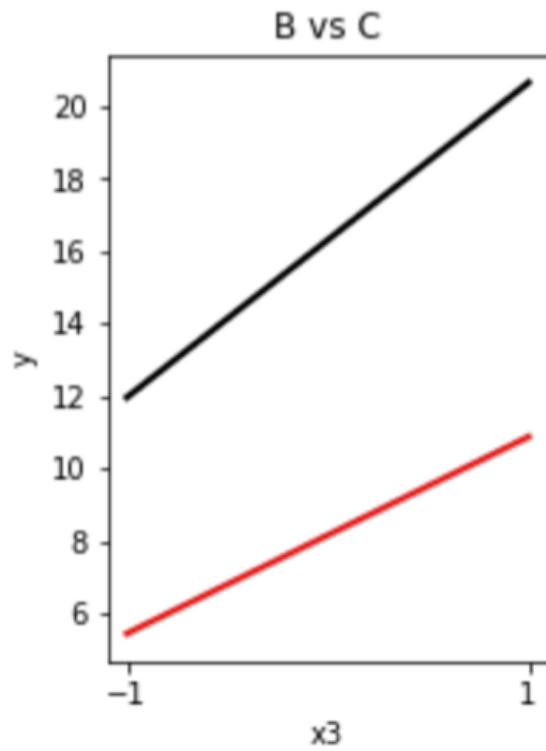
Python: regresjon

	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

Samspillsplott

$x_2 = +1$

$x_2 = -1$



Tips!

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