

ISTx1001: Industriell statistikk

22. oktober 2024

Vi starter klokka 14:15

Timen blir tatt opp

Thea Bjørnland, Institutt for matematiske fag, NTNU

Er det ønskelig å ha undervisning på mandager 14:15 - 15:00 i stedet for 15:15 - 16:00?

41

Ja

32 responses

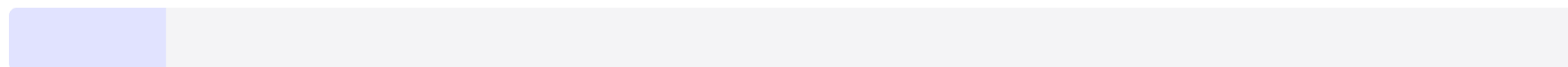
78%



Nei

4 responses

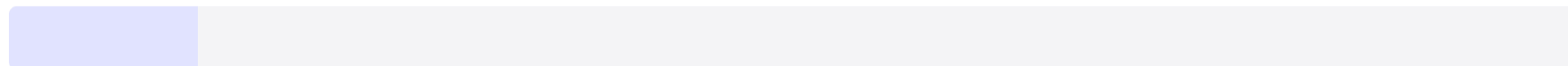
10%



Spiller ingen rolle

5 responses

12%



Undervisning (med opptak)

Mandager 14:15 - 15:00 (tre ganger til)

Tirsdager 14:15 - 16:00 (to ganger, inkludert i dag)

I dag: Steg 1, 2, 3, 4 i Del 1 av prosjektet

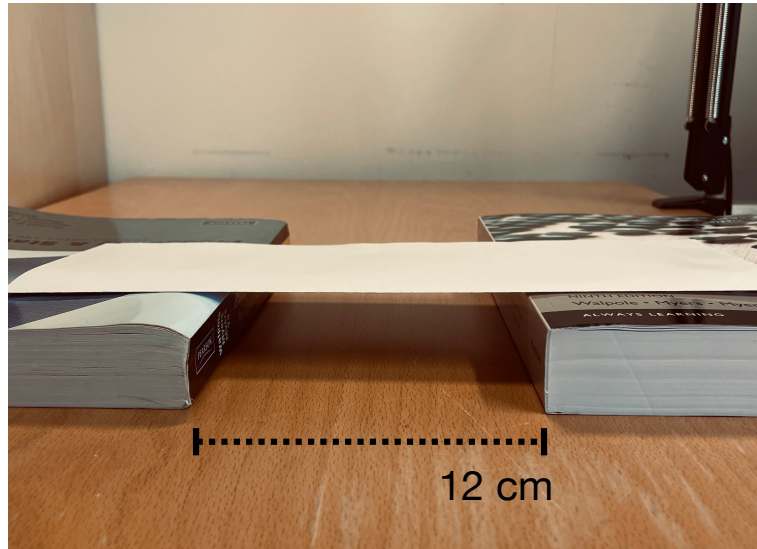
+ sammenhengen med lineær regresjon (del 5)

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?



Responsvariabel Y : bæreevne (g)

Måleinstrument/måte: 1-kr mynt
(4.35 g)

∴

Faktor: papirtype

↳ 2 nivåer

lavt nivå: grått ark, $x = -1$

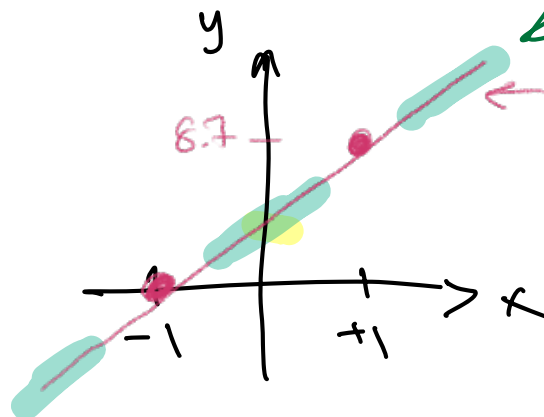
høyt nivå: hvitt ark, $x = +1$



4.35 gram

Data

x	y
-1	0
+1	8.7 g



← estimert regresjonslinje
 $\hat{\beta}_0 + \hat{\beta}_1 x = 4.35 + 4.35x$

obs: vi skal verken ekstrapolere eller
intrapolere, bare
interessert i
effekt for
 $x = -1$ og $x = +1$

Signifikant effekt eller tilfeldig variasjon?

Testobservator for $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$:

$$T = \frac{\hat{\beta}_1 - 0}{\text{SE}(\hat{\beta}_1)} \stackrel{H_0}{\sim} t_{n-2}$$

$\hat{\beta}_1 = 4.35$

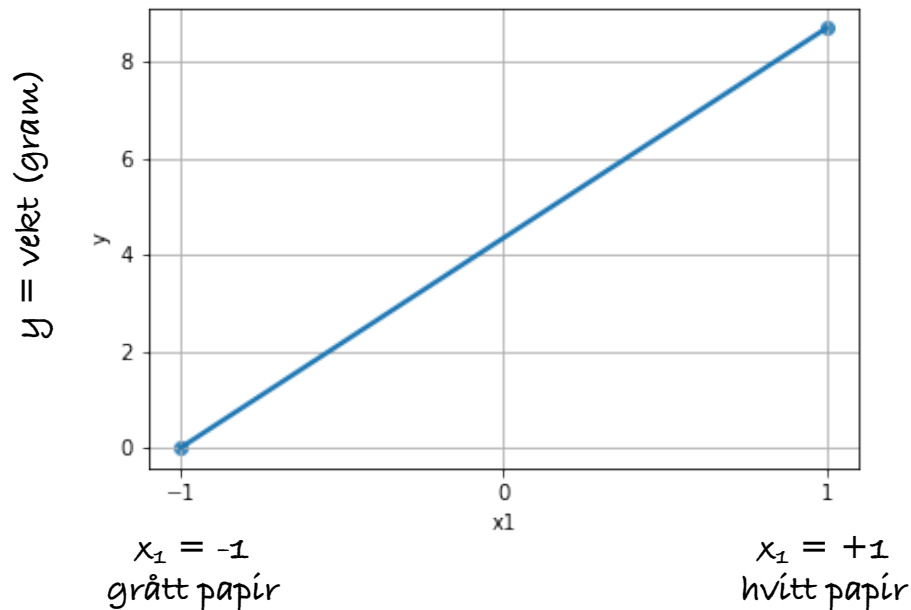
kan vi si med sikkerhet at bæreevnen endrer seg mellom de to faktornivåene?

$$\text{SE}(\hat{\beta}_1) = \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$n=2$
 $n-2=0$
 Δ i eksempelet vårt $S^2 = \frac{0}{0}$ (udefinert)

Signifikant effekt eller tilfeldig variasjon?



vi har brukt opp
de to datapunktene
til å estimere to parametre,
 β_0 og β_1

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.3500	inf	0	nan	nan	nan
x1	4.3500	inf	0	nan	nan	nan

La oss gjøre gjentak!

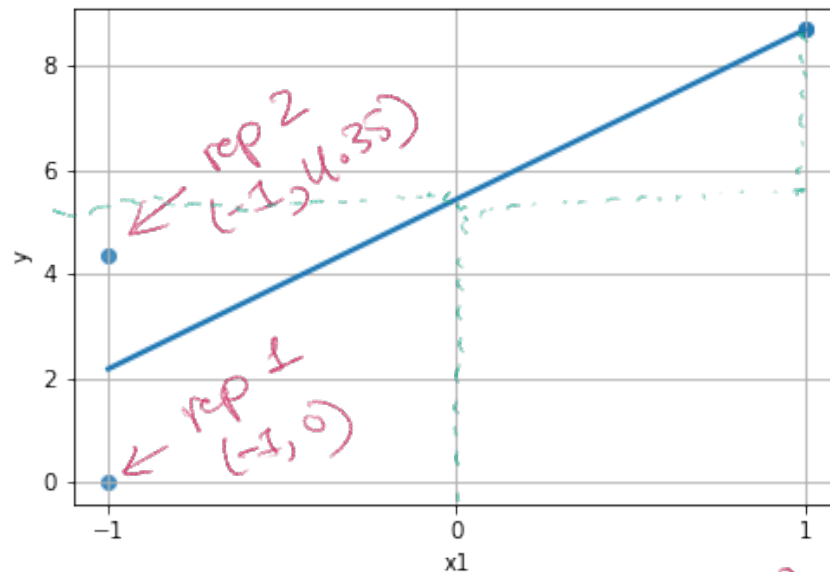
(og ha et forsøk der gjentakene ikke
alltid blir helt like)

Papirbro - flere målinger

rep 1

rep 2

x1	y
-1.0	0.00
1.0	8.70
-1.0	4.35
1.0	8.70



$$y = 5.44 + 3.2625 x_1$$

	coef	std err	t	P> t	[0.025	0.975]
Intercept	5.4375	1.087	5.000	0.038	0.758	10.117
x1	3.2625	1.087	3.000	0.095	-1.417	7.942

Python-udskrift

$$H_0: \beta_1 = 0 \text{ mod } H_1: \beta_1 \neq 0$$

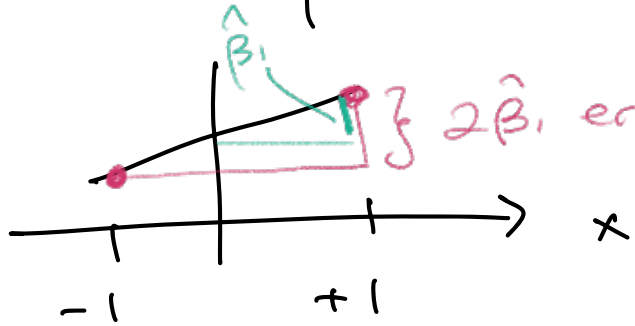
$$\text{Antagelse } Y \sim N(\beta_0 + \beta_1 x, \sigma) \rightarrow T\text{-test}$$

p-verdi fra t-test
= 0.095

Husk: minste kvadratsum gir

$$\hat{\beta}_1 = 3.26$$

(stigningstall til linja)



$2\hat{\beta}_1$ er estimert effekt av å bytte faktornivå (fra lavt til høyt nivå)

$$2\hat{\beta}_1 = 6.525g.$$

Dersom vi antar $Y = \beta_0 + \beta_1 x + \varepsilon$ $\varepsilon \sim N(0, \sigma)$

$Y|x \sim N(\beta_0 + \beta_1 x, \sigma)$ (ekvivalent)

Så kan vi regne ut KI og gjøre hypotesetest på β_1 ,

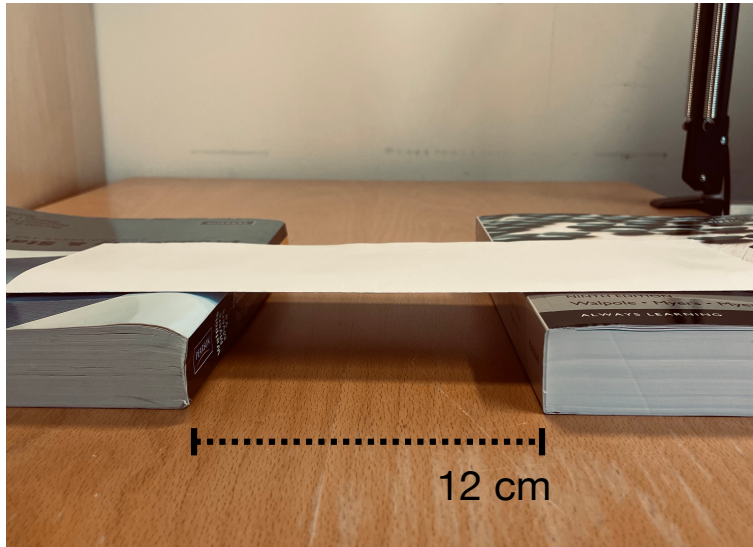
som går 95% KI $[-1.417, 7.942]$

dvs. stor usikkerhet i estimatet $\hat{\beta}_1$ og dermed stor usikkerhet i estimert effekt av faktor ($2\hat{\beta}_1$).

Kan ikke konkludere med at hvitt ark er bedre enn grønt.

Flere faktorer

2³ - forsøk
↑
3 faktorer
2 nivåer per faktor



A: papirtype (grått, hvitt)
B: avstand (8 cm, 12 cm)
C: bretteing (1 brett, 2 brett)

($x_1 = -1$, $x_1 = +1$)

($x_2 = -1$, $x_2 = +1$)

($x_3 = -1$, $x_3 = +1$)

numeriske koding

Multipl linear regresjon (del 2)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad \varepsilon \sim N(0, \sigma)$$

↑ avstand

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$SD(Y) = \sigma$$

Tolkning β_2 :

Endring i $E(Y)$ når x_2 øker med en enhet
mens x_1 og x_3 holdes konstante

ikke så
interessant
i 2-nivå
forsk

Tolkning av $2\beta_2$:

Effekt av å endre faktor B (avstand)
fra lavt til høyt nivå, mens de andre
faktorene holdes konstante.

Multippel linear regresjon (Del 2)

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

← papir ← avstand ← brekting

$$+ \underline{\beta_{12}} x_1 x_2 + \underline{\beta_{23}} x_2 x_3 + \underline{\beta_{13}} x_1 x_3$$

$$+ \underline{\beta_{123}} x_1 x_2 x_3 + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

— 2-faktor interaksjon

~ 3-faktor interaksjon ←

tar gjerne
bort denne.

NB: 8 koeffisienter i $E(Y)$ + σ i $\epsilon \sim N(0, \sigma^2)$

Lineær regresjon på et
uhensiktsmessig datasett

Eks: løpedata

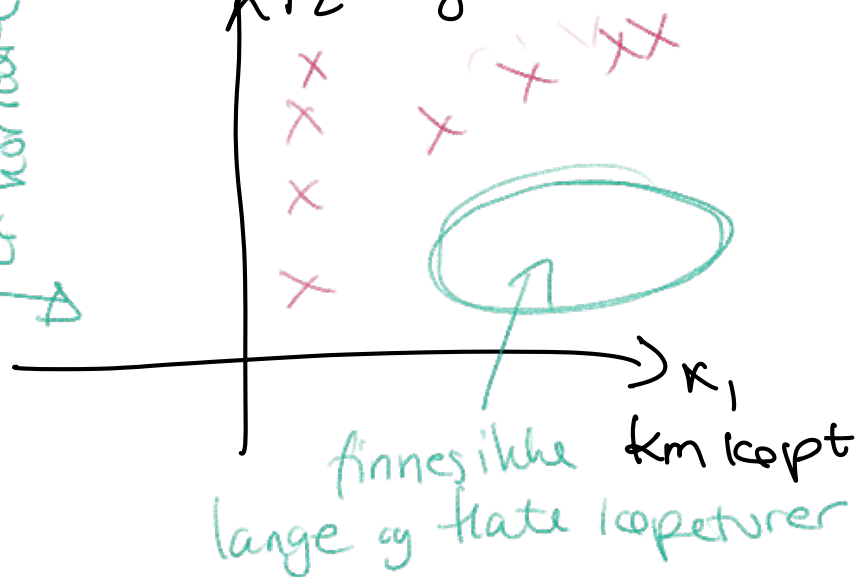
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

↑
kaloriforbruk
↑
km løpt
↑
høyde-
meter

(sjekk selv med notatbok)

Kovariater x_1 og x_2
↑
 x_2 høydemeter løpt

Obs: x_1 og x_2
er korrelerte



Lineær regresjon i en
designet studie.

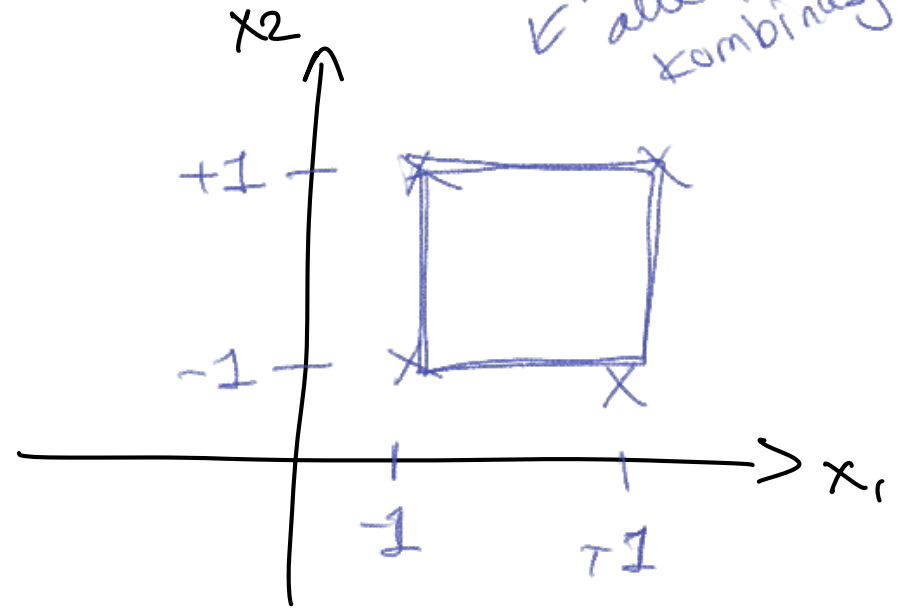
Eks: Papir bro

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

↑
bæreevne

← Papir
← avstand mellom bærer

← vi skal studere
alle faktorkombinasjoner



2-faktorer → $2 \cdot 2$ kombinasjoner
3-faktorer → 2^3 kombinasjoner

Designmatrise for 2^3 -forsøk

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

NB 1: Mønster

↳ må randomisere rekkefølgen på forsøkene når vi gjør data-innsamling.

NB 2: Med en repetisjon av forsøket får jeg 8 datapunkter

her er $k=3 \rightarrow 2^3=8$
bruk gjerne $k=4 \rightarrow 2^4=16$

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. $\rightarrow$ i les her blir Y diskret
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.

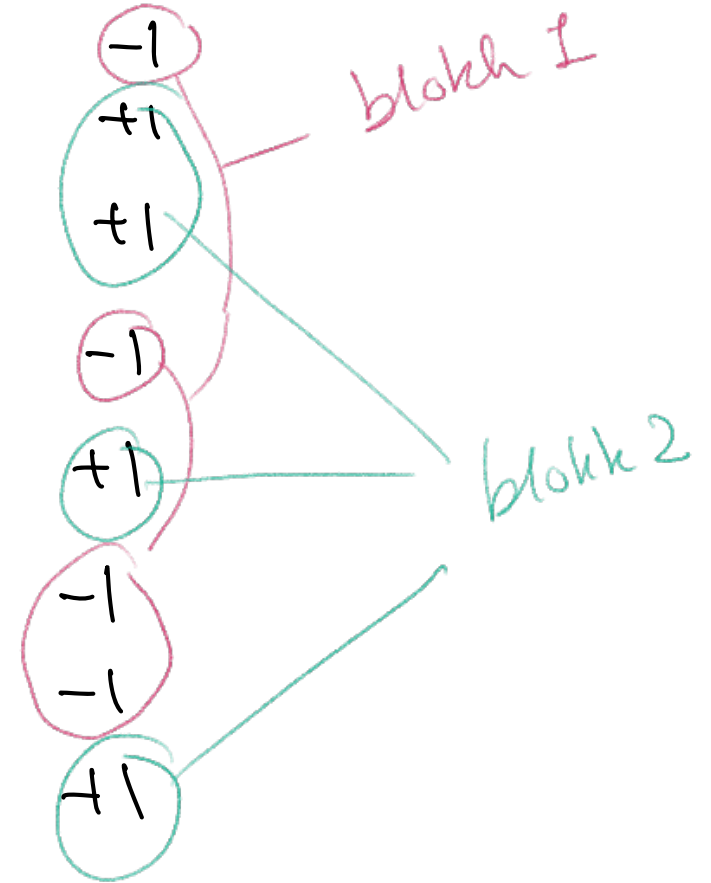
$2^3 \rightarrow$ må ha minst
ett gjentak
(2 repetisjoner)

Var realistisk,
kan gjerne
teste først?

Designmatrise → blokk inndeling (video 5)

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

triks
 $x_1 x_2 x_3$



- Bestem forsøk

↓ ↳ sjekk med følgerer f.eks forslag

+ designmatrise, randomisering,
antall gjentak , blokkdeling,
(regresjonsmodell)