

ISTx1001: Industriell statistikk

4. november 2024

Vi starter klokka 14:15

Timen blir tatt opp

Thea Bjørnland, Institutt for matematiske fag, NTNU

Prosjektarbeid i grupper

Uke 43, 44, 45, 46. Innlevering mandag uke 47

Undervisning

Del 1: Uke 43 og 44

Del 2/3: Uke 45 og 46

Veiledning

Alle uker

Innlevering

FØR
18. november
12:00 i INSPIRA

Plan

Forelesninger (små endringer kan forekomme)

- Mandag 21. oktober 15:15-16:00: Introduksjon
- Tirsdag 22. oktober 14:15-16:00: Forsøksplanlegging (del 1)
- Mandag 28. oktober 15:15-16:00: Forts. forsøksplanlegging (del 1)
- Tirsdag 29. oktober 14:15-16:00 Forts. forsøksplanlegging (del 1)
- **Mandag 4. november 14:15-15:00: Statistisk kvalitetsstyring (del 2)**
- Mandag 11. november 14:15-15:00: Statistisk kvalitetsstyring (del 2)

- Forelesningene vil i all hovedsak utdype prosjektoppgavene og svare på spørsmål. Pensum dekkes av kompendium, videoer og notatbøker.

Tips til data-analyse i del 1

istx100y2024 /

9/17/24, 3:08 PM	Last Modified
1001	18 seconds ago
1002	4 days ago
1003	16 days ago
forprosjekt	last month
uke1	last month
uke2	2 months ago
uke4	last month
uke5	last month
uke6	last month
uke7	27 days ago
uke8	last month
uke9	16 days ago
README.md	3 months ago
velkommenIS...	2 months ago

istx100y2024 / 1001 /

Name	Last Modified
DOE_EKS1.ipynb	6 days ago
DOE_EKS2.ipynb	6 days ago
DOE_EKS3.ipynb	10 days ago
DOE_introduk...	6 days ago
DOE_multipp...	17 days ago
DOEfunksjone...	30 seconds ago
papirbro.ipynb	17 days ago
produktunifor...	17 days ago

Tips til data-analyse i del 1

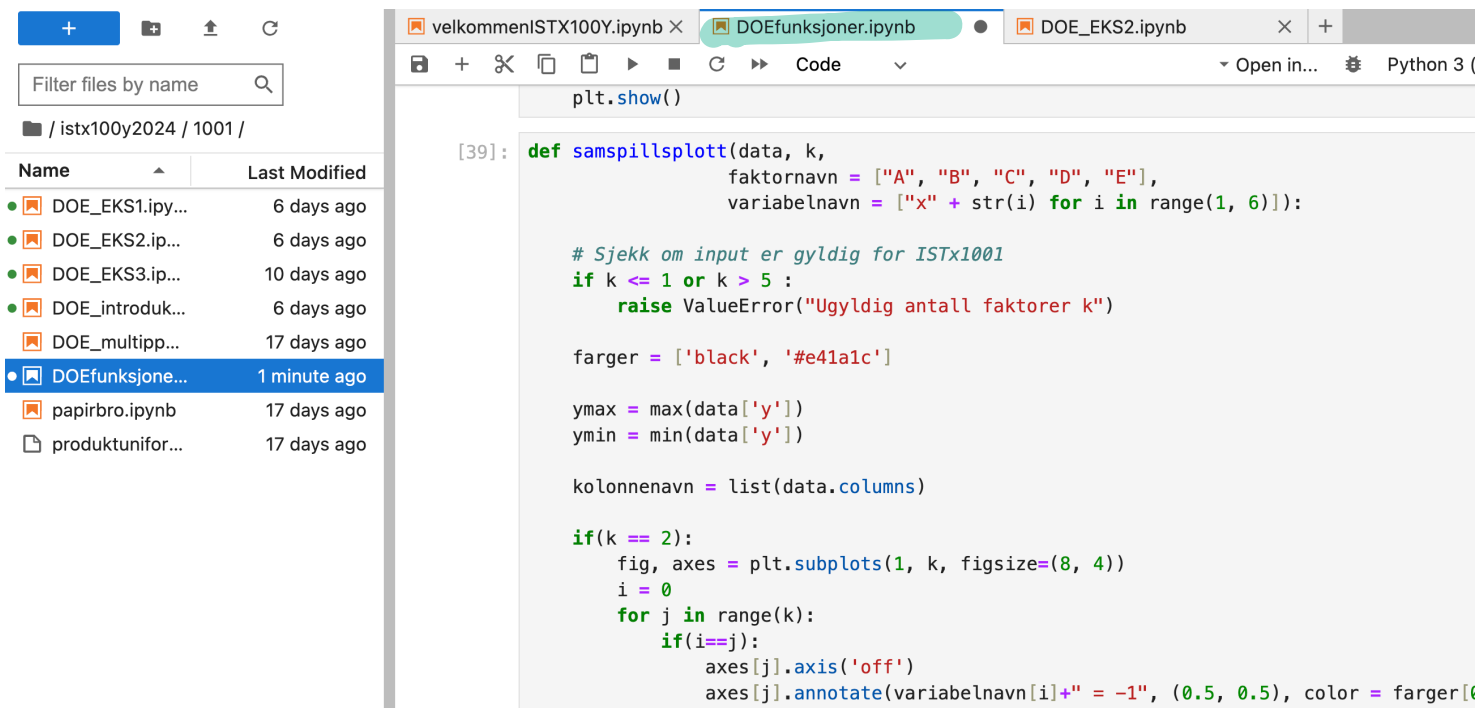
DOE eksempel 2: Produktuniformitet

Eksempelet og dataene er hentet fra National Institute of Standards and Technology (USA)

<https://www.itl.nist.gov/div898/handbook/pri/section3/pri3331.htm>

```
samspillsplott(data = df, k = 3)
```

→ 2^3 - forsøk



The screenshot shows a Jupyter Notebook interface. On the left is a file explorer for the directory `/istx100y2024 / 1001 /`. It contains a table of files:

Name	Last Modified
DOE_EKS1.ipynb	6 days ago
DOE_EKS2.ipynb	6 days ago
DOE_EKS3.ipynb	10 days ago
DOE_introduk...	6 days ago
DOE_multipp...	17 days ago
DOEfunksjone...	1 minute ago
papirbro.ipynb	17 days ago
produktunifor...	17 days ago

The main area shows a code cell with the following Python code:

```
[39]: def samspillsplott(data, k,
        faktornavn = ["A", "B", "C", "D", "E"],
        variabelnavn = ["x" + str(i) for i in range(1, 6)]):

    # Sjekk om input er gyldig for ISTx1001
    if k <= 1 or k > 5 :
        raise ValueError("Ugyldig antall faktorer k")

    farger = ['black', '#e41a1c']

    ymax = max(data['y'])
    ymin = min(data['y'])

    kolonnenavn = list(data.columns)

    if(k == 2):
        fig, axes = plt.subplots(1, k, figsize=(8, 4))
        i = 0
        for j in range(k):
            if(i==j):
                axes[j].axis('off')
                axes[j].annotate(variabelnavn[i]+" = -1", (0.5, 0.5), color = farger[i])
```

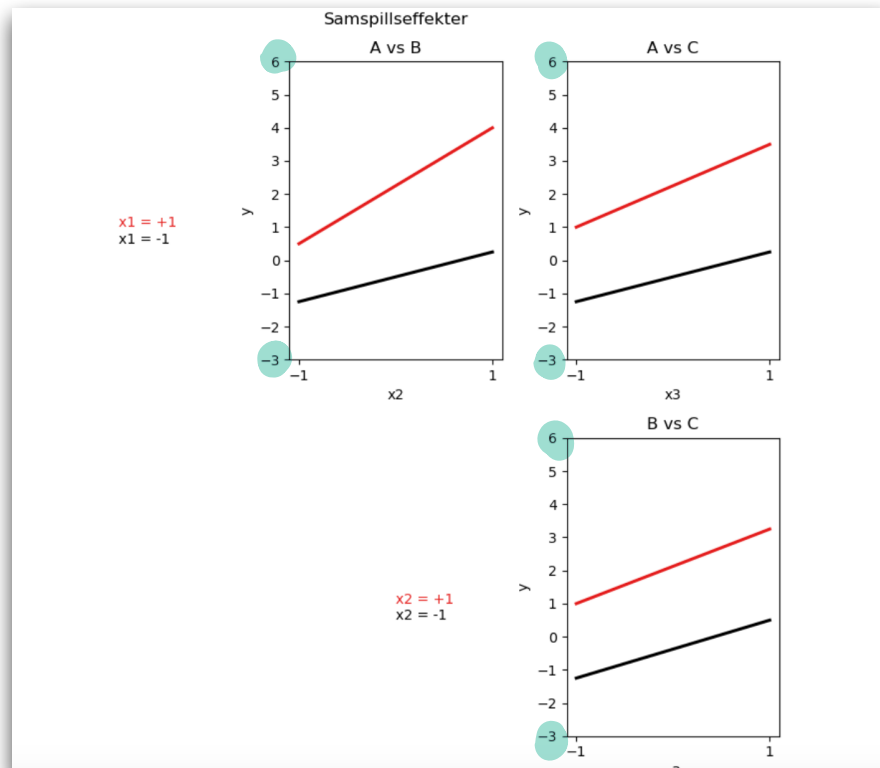
Tips til data-analyse i del 1

DOE eksempel 2: Produktuniformitet

Eksempelet og dataene er hentet fra National Institute of Standards and Technology (USA)

<https://www.itl.nist.gov/div898/handbook/pri/section3/pri3331.htm>

```
samspillsplott(data = df, k = 3)
```



Digital veiledning i morgen

<https://wiki.math.ntnu.no/istx100y/2024h/1001#veiledning>

Veiledning

Digitalt: Bruk [mattelab-forumet vårt](#) og underforumet 1001.

Zoom-veiledning Tirsdag 5. nov klokka 13:00 - 15:00, registrer tidspunkt i [dette skjemaet](#), zoom-lenke i skjemaet.

	A	B	C	D	E	F
1	Digital veiledning 5. november 2024 ISTx1001 prosjekt					
2	TID	NAVN (Dette navnet må dere også ha i zoom venterom)				
3	13:00					
4	13:10					
5	13:20					
6	13:30	NAVN				
7	13:40					
8	13:50					
9	Pause					
10	14:10					
11	14:20					
12	14:30					
13	14:40					
14	14:50					
15						
16						
17	Zoom:	https://NTNU.zoom.us/j/95727089069?pwd=EbbcQvSNw4nvqAFveM9r1gFXy4941K.1				
18						

Del 2 av prosjektet i ISTx1001

Ressurser:

Læreboka kapittel 11

Temavideoer (laget av Ketil Arnesen)

Forelesninger med slides

Del 2 av prosjektet i ISTx1001

Oppgave 2 - Statistisk prosesskontroll

Målet med oppgaven er å bestemme kontrollgrenser i en prosesskontroll, samt visualisere og forklare hva dette innebærer ved hjelp av stokastiske simuleringer.

Tenk dere at dere skal lage et produkt basert på forsøket dere gjennomførte i Oppgave 1, f.eks muffins. Dere må selv velge nivåer for forklaringsvariablene som produktet skal produseres under. En stor produksjon skal settes i gang, og responsvariabelen fra Oppgave 1 skal brukes for å kontrollere (med stikkprøver) at prosessen er under kontroll.

papirbroa: hvitark, 8cm, 2bredd. (den beste)

Python

Hente ut
data fra del 1.

```
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

data fra
del 1.

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

```
modell.predict()
```

```
array([ 5.709375, 18.215625,  2.990625,  7.884375,  9.515625, 31.809375,  
        3.534375, 18.215625,  5.709375, 18.215625,  2.990625,  7.884375,  
        9.515625, 31.809375,  3.534375, 18.215625])
```

$\hat{y} = 31,81$ gram

når $x_1 = 1$
 $x_2 = -1$
 $x_3 = 1$

forventet
bæreevne for
valgt innstilling (hvit, 8cm, 2 brett)

Python

```
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

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

```
modell.mse_resid
```

8.541406249999998

estimeret
varians til
regresjon

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

Standardavvik = $\sqrt{8.54...} \approx 2.92 \text{ gram.}$

$\bar{Y}$ = beregne brø med valgt instilling

$\bar{Y} \sim N(31.81, 2.92)$

[antagelse, estimerte
tall for μ og σ , men
antar dette som
en 'sannhet' for del 2.]

Stokastisk simulering: normal prosess

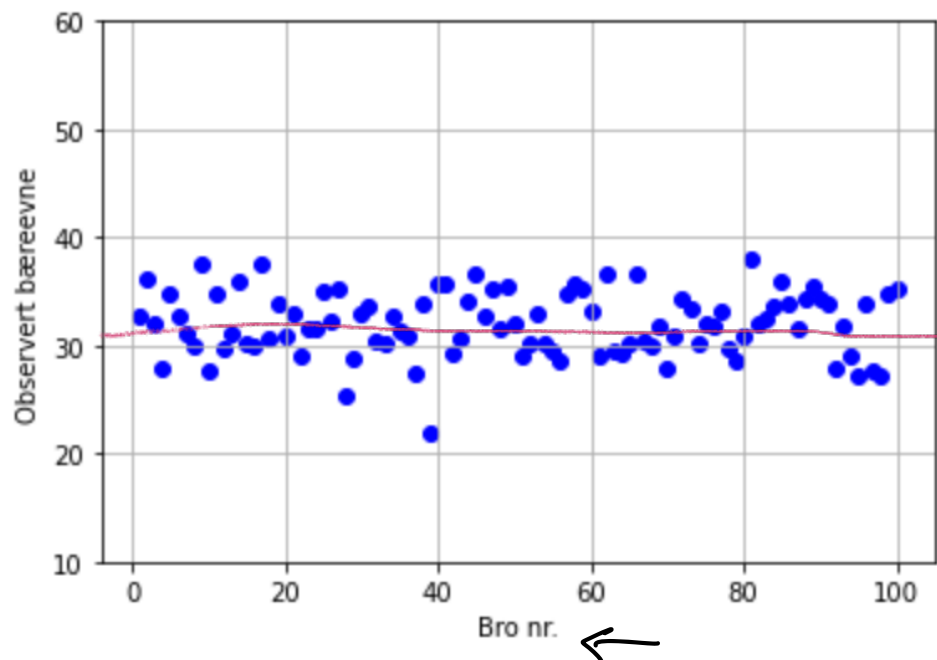
$\mu \rightarrow$ mean = 31.81
 $\sigma \rightarrow$ std_dev = 2.92
antall_broer = 100
bæreevne = np.zeros(antall_broer)

$\leftarrow 31.81$ $\leftarrow 2.92$

```
for i in range(antall_broer):  
    bæreevne[i] = np.random.normal(mean, std_dev)
```

bro_nr = np.arange(1, antall_broer + 1)

```
plt.scatter(bro_nr, bæreevne, marker='o', color='b')  
plt.ylim(10,60); plt.xlabel("Bro nr."); plt.ylabel("Observert bæreevne")  
plt.grid()  
plt.show()
```



$\leftarrow$ variasjon skyldes
normalfordeling med
standardavvik $\sigma = 2.92$

$\mu = 31.81$

Stokastisk simulering: hva kan gå galt?

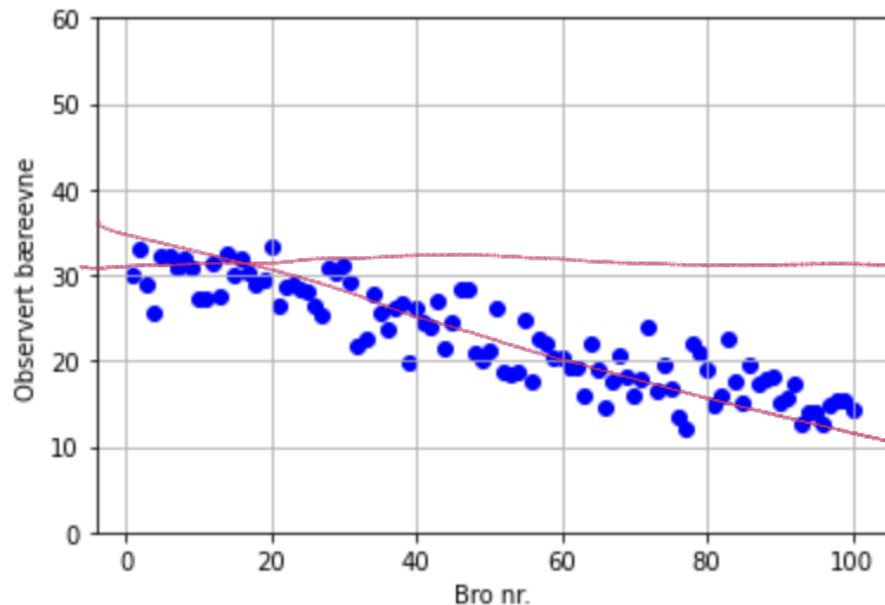
```
mean = 31.81
std_dev = 2.92
antall_broer = 100
bæreevne = np.zeros(antall_broer)

for i in range(antall_broer):
    if (i + 1) % 5 == 0:
        mean -= 0.3 * std_dev
        bæreevne[i] = np.random.normal(mean, std_dev)

bro_nr = np.arange(1, antall_broer + 1)

plt.scatter(bro_nr, bæreevne, marker='o', color='b')
plt.ylim(0,60); plt.xlabel("Bro nr."); plt.ylabel("Observert bæreevne")
plt.grid()
plt.show()
```

← systematisk kvalitetsavvik
→ bæreevnen blir stadig
dårligere etter hvert som
flere broer produseres.



$\mu = 31.81$

forventningsverdien
synker.

Stokastisk simulering: hva kan gå galt?

μ
6

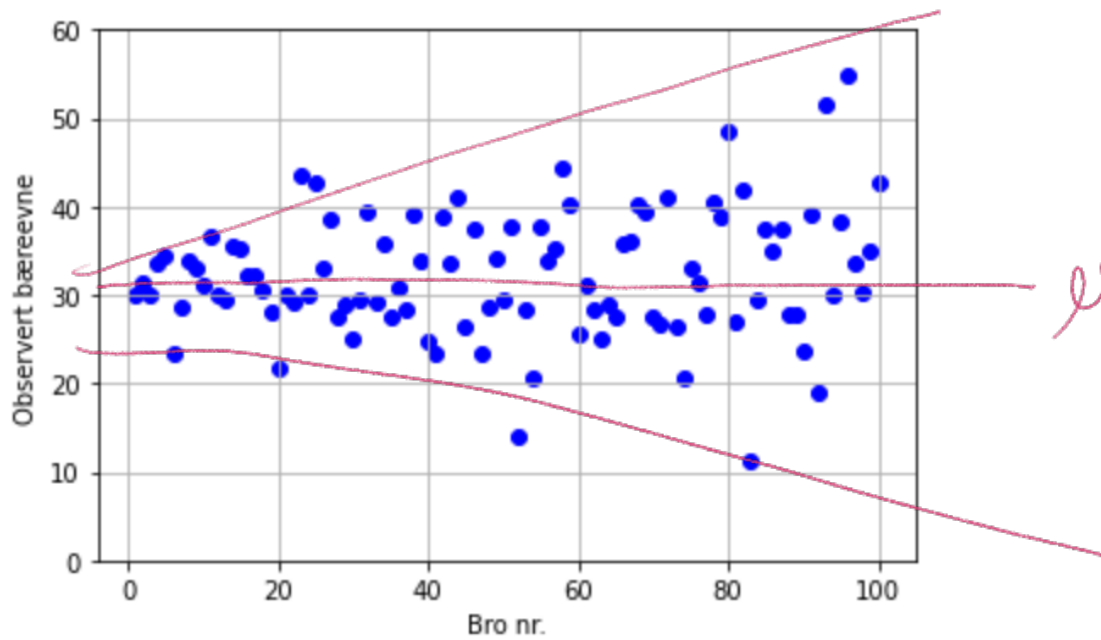
```
mean = 31.81
std_dev = 2.92
antall_broer = 100
bæreevne = np.zeros(antall_broer)

for i in range(antall_broer):
    if (i + 1) % 5 == 0:
        std_dev += 0.5
        bæreevne[i] = np.random.normal(mean, std_dev)

bro_nr = np.arange(1, antall_broer + 1)

plt.scatter(bro_nr, bæreevne, marker='o', color='b')
plt.ylim(0,60); plt.xlabel("Bro nr."); plt.ylabel("Observert bæreevne")
plt.grid()
plt.show()
```

Kvalitetsavvik
→ mer variasjon



$\mu = 31.81$

Stokastisk simulering: hva kan gå galt?

Vi må følge med på forventningsverdi og standardavvik!

* Vi ønsker å oppdage systematiske kvalitetsavvik i produksjon

* Vi vil ikke kontrollere alle produkter (dyrt)

↙
* Sjekk et (lite) tilfeldig utvalg } stikkprøve.
med jevne mellomrom

Eks: $n=5$, y_1, y_2, y_3, y_4, y_5

① Er $\bar{y}$ i nærheten av $\mu = 31.81$?

② Er s i nærheten av $\sigma = 2.92$?

← måling av bæreener
på 5 tilfeldig valgte
papirbroer.

} kontrollgrenser

Del 2 av prosjektet i ISTx1001

$\mu = 31.81$
 $\sigma = 2.92$

2a) Kontrollgrenser

Finn en estimert forventningsverdi og et standardavvik på Y fra Oppgave 1 for den valgte faktorkombinasjonen. Dette setter dere som verdier på μ og σ for en prosess som er under kontroll. Bestem hvor mange (n) produkter dere vil kontrollere i hver stikkprøve. $n = 5$

- ① Bestem kontrollgrenser for gjennomsnittet til en stikkprøve. Bruk Shewarts' anbefaling om 3-sigma-grenser.
- ② Bestem kontrollgrenser for det empiriske standardavviket til en stikkprøve basert på kritiske verdier fra kjikkvadratfordelingen med $\alpha/2 = 0.002$.

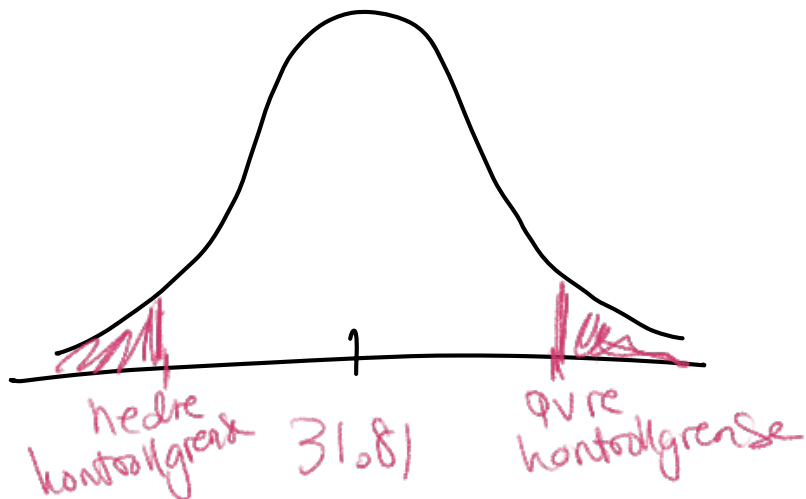
Kontrollgrenser for gjennomsnittet

H_0 / under kontroll

$$Y \sim N(\mu = 31,81, \sigma = 2,92)$$

Stikprøve $Y_1, Y_2, \dots, Y_5$

$$\bar{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$



Rep: hypotesetesting
• kritisk verdi
 α = signifikansnivå
 $u = Z_{\alpha/2}$
 $\alpha = 0,05$
 $Z_{\alpha/2} = 1,960$
→ 5% sjanse for falsk alarm.

Kontrollgrenser for gjennomsnittet

3-sigma-grenser

Øvre grense : $\mu + 3 \cdot \frac{\sigma}{\sqrt{n}}$ = 35.72 g

Handwritten notes:
31.81 (pointing to μ)
2.96 (pointing to 3)
1.5 (pointing to $\sqrt{n}$)
kritisk verdi ($z_{\alpha/2}$) (pointing to 3)

nedre grense : $\mu - 3 \cdot \frac{\sigma}{\sqrt{n}}$ = 27.89 g

$$P(\text{falsk alarm}) = \alpha = 0.0027$$

Kontrollgrenser for standardavviket

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

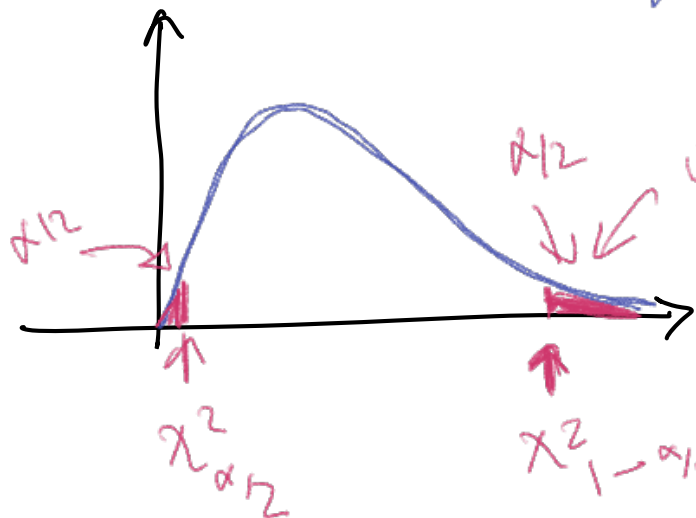
↑ estimator for variance

hva slags fordeling
har s^2 når alt
er under kontroll?

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

kji-kvadrat-fordeling
(chi-squared)

$n=5$



usannsynlig
når prosessen er
under kontroll,
peker på noe galt
med variasjon.

Kontrollgrenser for standardavviket

```
stats.chi2.ppf(0.002, 4)
```

→ 0.12923771486623198

```
stats.chi2.ppf(1-0.002, 4)
```

→ 16.923758195804652

$$\frac{\alpha}{2} = 0.002$$

$$n-1=4$$

nexte gang:

Øvre og nedre kontrollgrense for
S (empirisk standardavvik)

+ stokastisk simulering av stikprøver.

Del 2 av prosjektet i ISTx1001

2a) Kontrollgrenser

Finn en estimert forventningsverdi og et standardavvik på Y fra Oppgave 1 for den valgte faktorkombinasjonen. Dette setter dere som verdier på μ og σ for en prosess som er under kontroll. Bestem hvor mange (n) produkter dere vil kontrollere i hver stikkprøve.

Bestem kontrollgrenser for gjennomsnittet til en stikkprøve. Bruk Shewarts' anbefaling om 3-sigma-grenser.

Bestem kontrollgrenser for det empiriske standardavviket til en stikkprøve basert på kritiske verdier fra kjikvadratfordelingen med $\alpha/2 = 0.002$.