

# Opgave 1: Sandsynlighedsregning

Et system er udviklet til at registrere om RFID tags er defekte. En undersøgelse viser, at givet et RFID tag ikke er defekt, er sandsynligheden for, at systemet registrerer, at det ikke er defekt, lig med 0,4. Og givet et RFID tag er defekt, er sandsynligheden for, at systemet registrerer, at det er defekt, lig med 0,999.

**Vi ved at 1 ud af 100 RFID tags er defekte.**

Vi kender til disse værdier

$R$  = RFID IKKE DEFJEKT

$\text{Not}_R$  = RFID DEFJEKT

$S$  = System registrere ingen fejl

$\text{Not}_S$  = System registrere fejl

$$\Pr(S|R) = 0.4$$

$$\Pr(\overline{S} | \overline{R}) = 0.999$$

$$\Pr(\overline{R}) = \frac{1}{100} = 0.01$$

**a) Hvad er sandsynligheden for at et RFID tag ikke er defekt?**

$$\Pr(R) = 1 - \Pr(\overline{R}) = 0.99$$

0.99

**b) Hvad er den totale sandsynlighed for at systemet registrerer et RFID tag som defekt?**

$$\Pr(\overline{S}|R) = 1 - \Pr(S|R) = 0.6$$

$$\Pr(S) = \Pr(S|R) \cdot \Pr(R) + \Pr(\overline{S}|\overline{R}) \cdot \Pr(\overline{R}) = 0.396$$

$$\Pr(\overline{S}) = 1 - \Pr(S) = 0.604$$

Den totale sandsynlighed for at registrere defekt tag er 60.4%

**c) Hvad er sandsynligheden, givet at et RFID tag registreres som defekt, at det også er defekt?**

Bayes regel:

$$\Pr(R|\overline{S}) = \frac{\Pr(\overline{S}|R) \cdot \Pr(R)}{\Pr(\overline{S})} = 0.983$$

$$\Pr(\overline{R}|\overline{S}) = 1 - \Pr(R|\overline{S}) = 0.017$$

givet at et RFID tag registreres som defekt, er sandsynligheden for at det også er defekt 1.7%

**d) Hvad er sandsynligheden, givet at et RFID tag registreres som defekt, at det ikke er defekt?**

$$\Pr(R|\overline{S}) = 0.983$$

givet at et RFID tag registreres som defekt, er sandsynligheden for at det IKKE er defekt 98.3%

## Opgave 2: Stokastiske variable

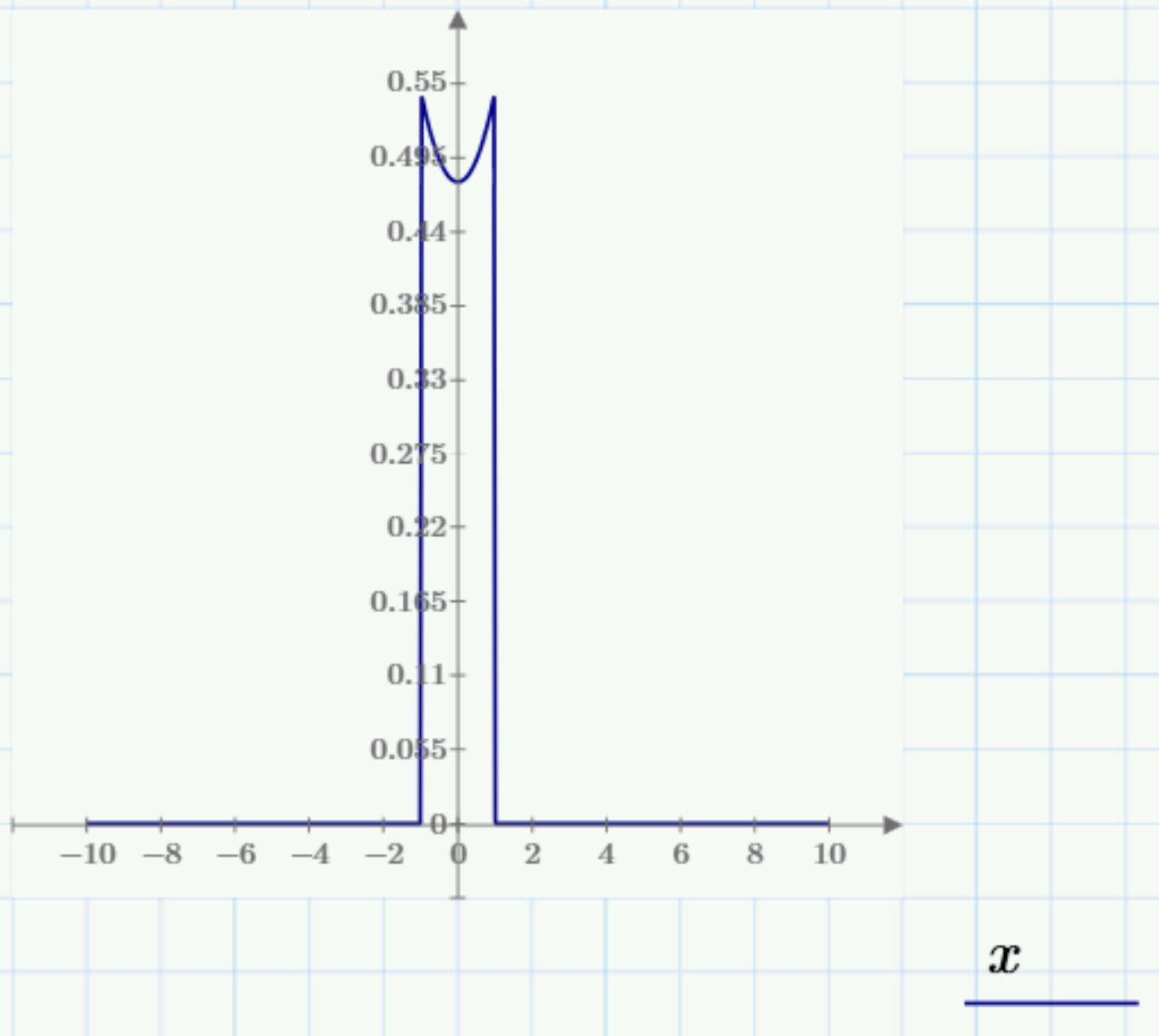
En tæthedsfunktion (pdf) for en stokastisk variabel er defineret til at være:

$$f_X(x) = \begin{cases} K \cdot (x^2 + 7) & \text{if } -1 < x < 1 \\ 0 & \text{else} \end{cases}$$

- a) Vis at  $K=3/44$ , for at  $f_X(x)$  er en gyldig tæthedsfunktion (pdf). Skitsér  $f_X(x)$

$$\int_{-1}^{1} K \cdot (x^2 + 7) dx = 1 \xrightarrow{\text{solve, } K} \frac{3}{44}$$

$f_X(x)$



b) Opstil udtrykket for og find middelværdien (forventningsværdien) af  $X$

middelværdi:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)$$

$$EX := \int_{-\infty}^{\infty} x \cdot f_X(x) dx = 0$$

Middelværdi er 0

**varians:**

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx$$

$$\text{Var}(x) = \sigma_X^2 = E[X^2] - E[X]^2 = 0.345 - 0^2 = 0.345$$

$$EX2 := \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx = 0.345$$

$$var := EX2 - EX^2 = 0.345$$

variansten er 0.345

c) Opstil udtrykket for og find standard afvigelsen af X.

$$SD(X) = \sigma_X = \sqrt{\sigma_X^2}$$

$$SD := \sqrt{var} = 0.588$$

standard afvigelsen er 0.588

d) Opstil udtrykket for og find fordelingsfunktionen (cdf'en) for X.

$$Fx(x) := \int_{-1}^x \frac{3}{44} \cdot (x^2 + 7) dx \rightarrow \frac{x^3}{44} + \frac{21 \cdot x}{44} + \frac{1}{2}$$

$$F_x(x) = \begin{cases} 0 & , \text{for } x < -1 \\ \frac{x^3}{44} + \frac{21 \cdot x}{44} + \frac{1}{2} & , \text{for } -1 \leq x < 1 \\ 1 & , \text{for } x > 1 \end{cases}$$

## Opgave 3: Stokastiske processer

En kontinuert stokastisk proces er givet ved:

$$y(t) = x(t) + w$$

hvor  $x(t) \sim \mathcal{N}(0, t^2)$  og  $w \sim \mathcal{N}(3, 1)$ , og  $x$  og  $w$  er uafhængige.

a) Tegn 3 realisationer af  $y$ , der er samplet med en periodetid på 1s, i intervallet

$0s \leq t \leq 10s$ . Skriv hvorledes realisationerne er fremkommet ved brug af en Gauss generator, f.eks. `randn()` i Matlab.

```
%Three realizations 11 samples

t=0:1:10;
xmu = 0;
wsigma=1;
wmu=3;
T=11;
for i=1:3 %1:3 because we need 3 realizations
x=(randn(1,11).*t);
w=(randn+3);
yt=x+w
figure;
n=1:10;%10 samples between 10-1
plot(t,yt,'kx')
grid
```

```

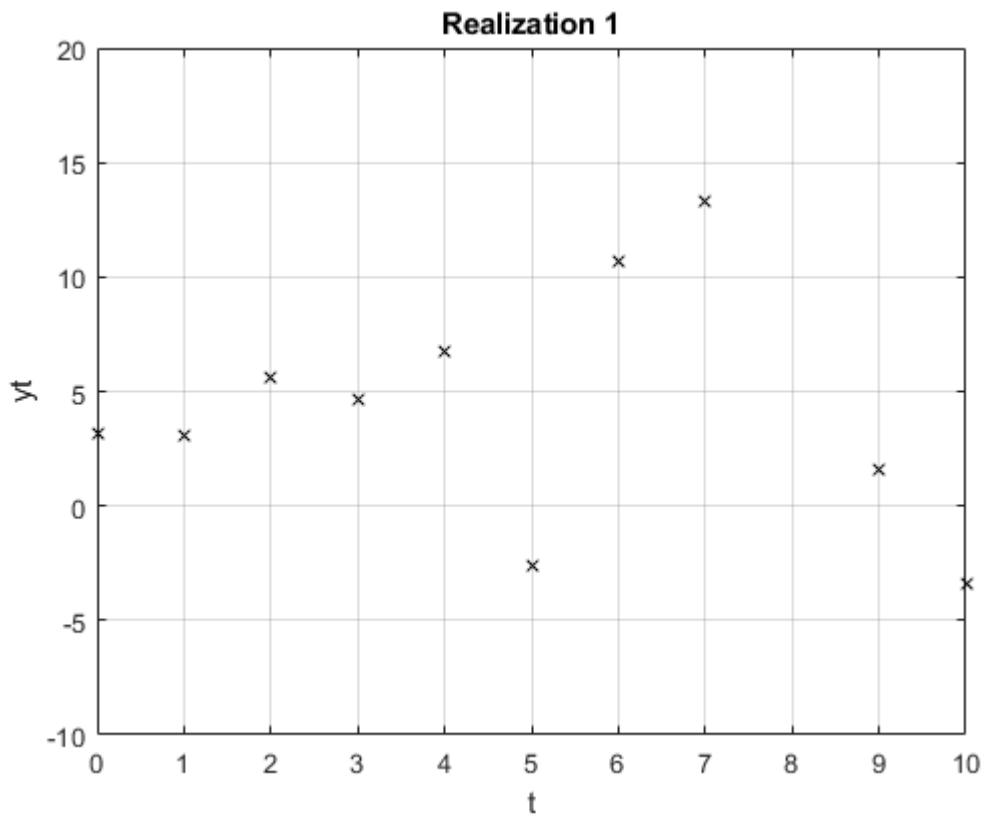
axis([0,10,-10,20])%set values(length) for x and y axis
title(['Realization ',num2str(i)])%axis title
xlabel('t');
ylabel('yt');
end

```

```

yt = 1x11
3.1565    3.0396    5.6090    4.6414    6.7909   -2.6146   10.7019   13.3243 ...

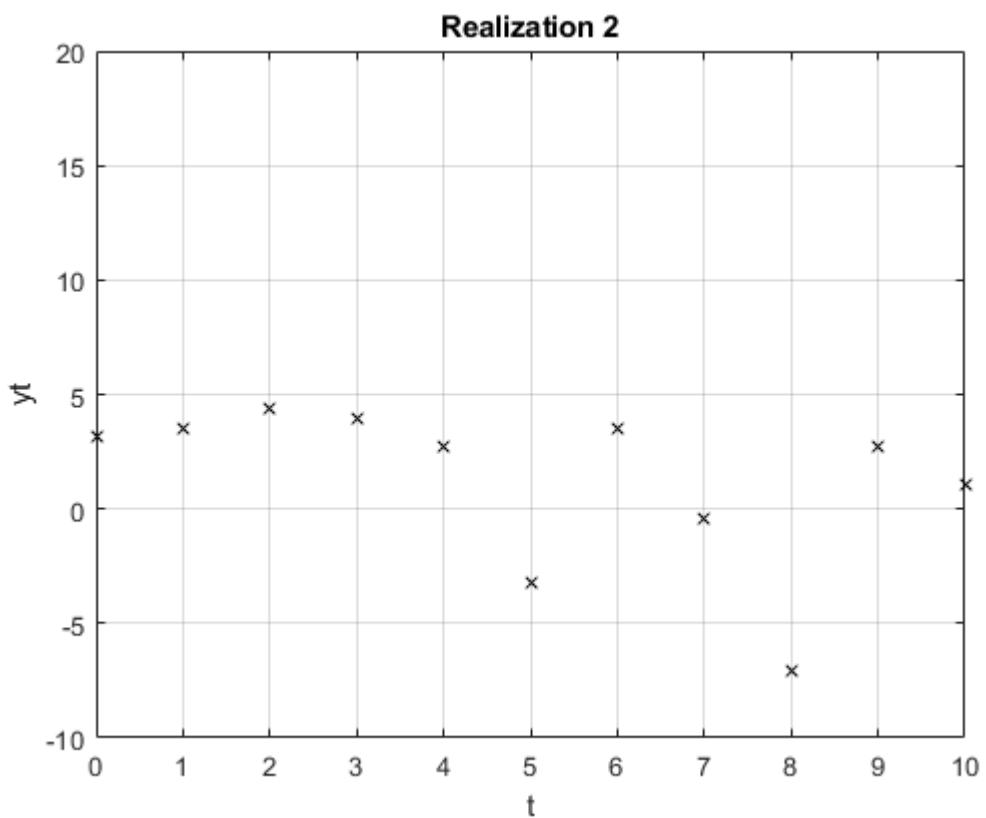
```



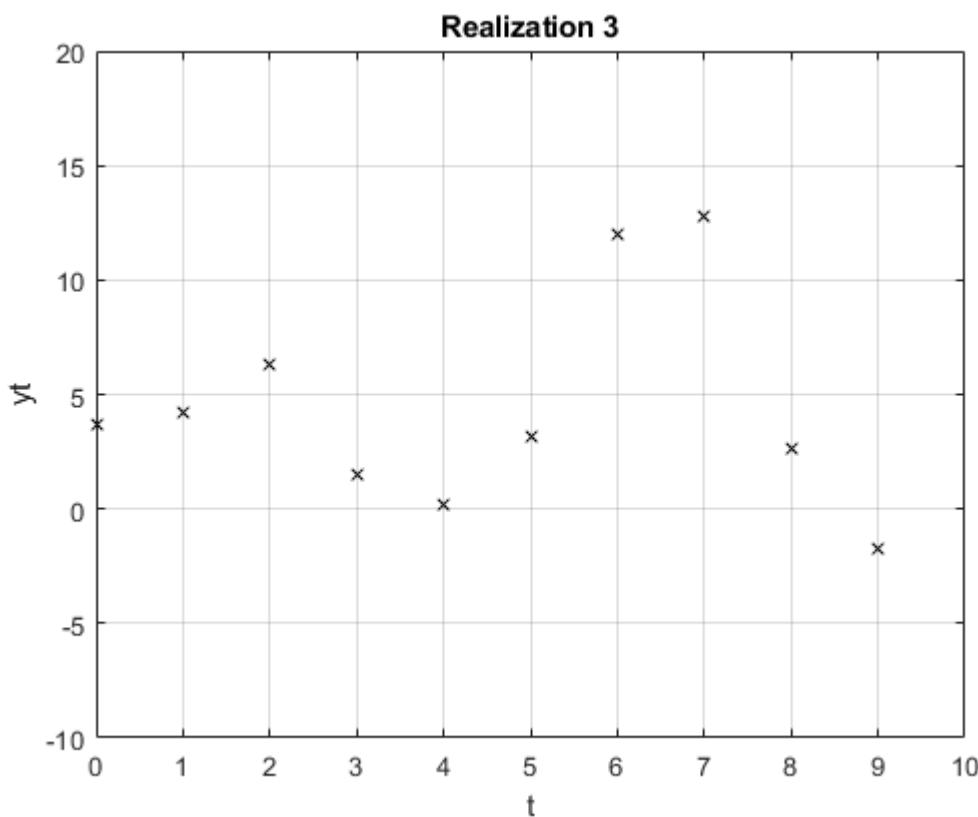
```

yt = 1x11
3.1921    3.4711    4.3863    3.9263    2.7187   -3.2549    3.5375   -0.4544 ...

```



```
yt = 1x11
    3.6718    4.2481    6.2834    1.4839    0.2135    3.1972    11.9711    12.7976 ...
```



b) Opskriv et udtryk og udregn ensemble middelværdien for processen  $y(t)$ .

middelværdi for kontinueret uniform er  $\mu$

så formlen ser således ud

$$E[yt] = E[x] + E[W] = 0 + 3 = 3$$

```
ensemble_mean=0+3
```

```
ensemble_mean = 3
```

ensemble mean: 3

c) Opskriv et udtryk og udregn ensemble variansen for processen  $y(t)$ .

Variansen for en kontinueret uniformfordeling er  $\sigma^2$

$$\text{var}(t) = \text{var}[x] + \text{var}[w] = t^2 + 1 = t^2 + 1$$

ensemble var:  $t^2 + 1$

d) Opskriv udtryk for den tidslige middelværdi og varians for én realisation af processen  $y(t)$ .

$$\text{temporalMiddel} = \hat{\mu}_{X_i} = \lim_{T \rightarrow \infty} \cdot \frac{1}{T} \cdot \sum_{i=0}^n y_i(t)$$

$$\text{temporalVarians} = \hat{\sigma}_{X_i}^2 = \lim_{T \rightarrow \infty} \cdot \frac{1}{T} \cdot \sum_{i=0}^n (y_i(t)^2 - \hat{\mu}_{X_i}^2)$$

```
temporal_mean=(1/T)*sum(x+w)
```

```
temporal_mean = 3.0270
```

```
temporal_var=vpa((1/T)*sum(((x+w).^2))-w.^2)
```

```
temporal_var = 35.027411896066453778075810987502
```

e) Er processen  $y(t)$  stationær i den brede forstand (WSS)? Er den ergodisk? Begrund dine svar.

Den er ikke WSS eller ergodisk.

Tid er taget i betragtning derfor er den ikke WSS

Den er ikke ergodisk fordi dens temporale og ensemble varians, samt middelværdi ikke er ens.

## Opgave 4: Statistik

Til kvalitetssikring af en produktion er udviklet et automatisk overvågningssystem, der skal registrere om en komponent er monteret korrekt. Overvågningssystemet skal i 95% af tilfældene registrere komponenterne korrekt, dvs. fejl-monterede komponenter registreres som fejl og korrekt-monterede komponenter registreres som korrekte.

For at undersøge om systemet overholder dette krav laves en række måleserier på et produkt A, hvor der testes om overvågningssystemets succesrate er præcis 95%. Testen består af i alt 15 måleserier. Hver serie består af 100 enheder, hvor antallet af korrekte registreringer (succeser) noteres.

## Testdata:

Måleserie A	1	2	3	4	5	6	
Antal korrekte registreringer	98	98	93	96	95	92	90

a) Bestem fra testen den estimerede succesrate (sandsynligheden for korrekt registrering).

```
AKR=[98 98 93 96 95 92 99 95 95 94 98 93 99 94 98];
trials=15*100
```

```
trials = 1500
```

```
succesiveTrials=sum(AKR)
```

```
succesiveTrials = 1437
```

```
sucessrate=succesiveTrials/trials
```

```
sucessrate = 0.9580
```

Der er en sucessrate på 0.958

b) Hvilken test-statistik kan bruges til at teste om overvågningssystemet overholder kravet til korrekt registrering. Der skal gøre rede for, at betingelserne for den benyttede test-statistik er opfyldt, samt hvilken NULL-hypotese og alternative hypotese, der opstilles til testen.

Binomialfordelings test med normal approximation - da der er tale om succes og fejl scenarier.

Vi tester først betingelserne

$$n \cdot p \Rightarrow 1500 \cdot 0.958 = 1437$$

```
trials*sucessrate
```

```
ans = 1437
```

$$n(1 - p) \Rightarrow 1500 \cdot (1 - 0.958) = 63.00$$

```
trials*(1-0.958)
```

```
ans = 63.0000
```

I det at begge betingelser er over værdien 5 kan vi antage at der er approximatalt er tale om en normalfordeling med  $\mu = 0$  og  $\sigma^2 = 1$

Vi kan nu opstille en nullhypotese

$H_0 : p_0 = 0.95$ , 95% af komponenterne registreres som forventet

$H_1 : p_0 \neq 0.95$ , 95% af komponenterne registreres IKKE som forventet

p0=0.95;

**c) Bestem p-værdien for de målte testdata. Opfylder overvågningssystemet kravet med et signifikans-niveau på 5%.**

Da dette er en binomial test kan vi se det som en normal distribution med  $\mu = 0$  og  $\sigma^2 = 1$

$$z = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}}$$

z=(succesiveTrials-(trials\*p0))/(sqrt((trials\*p0)\*(1-p0)))

z = 1.4216

$$p = 2 \cdot |1 - \Phi(|z|)|$$

p=2\*abs(1-normcdf(abs(z)))

p = 0.1551

I det at vores p-værdi er på 0.1551, hvilket er over 0.05. kan vi IKKE forkaste vores nullhypotese

**d) Bestem 95% konfidensintervallet for overvågningssystemet med de målte testdata.**

$$p_- = \frac{1}{n + 1.96^2} \left[ x + \frac{1}{n + 1.96^2} - 1.96 \sqrt{\frac{x(n-x)}{n} + \frac{1.96^2}{4}} \right]$$

$$p_+ = \frac{1}{n + 1.96^2} \left[ x + \frac{1}{n + 1.96^2} + 1.96 \sqrt{\frac{x(n-x)}{n} + \frac{1.96^2}{4}} \right]$$

Pmin=((1)/(trials+(1.96^2)))\*(succesiveTrials+((1)/(trials+(1.96^2)))-1.96\*sqrt(((succesiveTri

Pmin = 0.9453

Pmax=((1)/(trials+(1.96^2)))\*(succesiveTrials+((1)/(trials+(1.96^2)))+1.96\*sqrt(((succesiveTri

Pmax = 0.9658

Konfidensintervallet ligger mellem [0.9453;0.9658]

Da 0.95 indegår i konfidens intervallet kan vi ikke afvise vores nullhypotese

## Testdata:

Måleserie B	1	2	3	4	5	6	7
Antal korrekte registreringer	99	94	89	92	90	91	95

e) Hvilken test skal laves for at undersøge om de to datasæt er forskellige?

Uparret t-test Fordi vi har 2 test sæt som ikke er ens og har en ukendt varians.

f) Opstil NULL-hypotese, beregn p-værdien og konkluder om de to datasæt er ens med 5% signifikansniveau.

$$H_0 : \delta_0 = 0, \text{ datasættene er ens}$$

$$H_1 : \delta_0 \neq 0, \text{ datasættene er IKKE ens}$$

```
AKRB=[99 94 89 92 90 91 92 96 92 90 90 95];
trialsB=12*100
```

```
trialsB = 1200
```

```
succesiveTrialsB=sum(AKR)
```

```
succesiveTrialsB = 1110
```

```
succesRateB=succesiveTrialsB/trialsB
```

```
succesRateB = 0.9250
```

```
korrektMaalingPr100A=succesiveTrials/length(AKR)
```

```
korrektMaalingPr100A = 95.8000
```

```
korrektMaalingPr100B=succesiveTrialsB/length(AKR)
```

```
korrektMaalingPr100B = 92.5000
```

```
d=korrektMaalingPr100A-korrektMaalingPr100B
```

```
d = 3.3000
```

```
SampleVarianceA=var(AKR)
```

```
SampleVarianceA = 5.6000
```

```
SampleVarianceB=var(AKR)
```

```
SampleVarianceB = 8.8182
```

```
pooledVar=(1/((length(AKR)-1)+(length(AKRB)-1)))*((length(AKR)-1)*SampleVarianceA+(length(AKRB)
```

```
pooledVar = 7.0160
```

$$t = \frac{d}{\sqrt{\text{pooledVar}} \cdot \sqrt{\frac{1}{\text{length(AKR)}} + \frac{1}{\text{length(AKRB)}}}}$$

```
t=d/((sqrt(pooledVar))*(sqrt((1/length(AKR))+(1/length(AKRB)))))
```

```
t = 3.4119
```

```
pval = 2 · (1 - tcdf(|t|, n - 1))
```

```
p=2*(1-tcdf(abs(t),length(AKR)+length(AKRB)-2))
```

```
p = 0.0022
```

Da pværdien er under 0.05 kan vi forkaste hypotesen om at datasættene er ens, hvilket betyder at de 2 systemer virker forskelligt