E2015

Opgave 1: Stokastiske Variable

Lad den simultane tæthedsfunktion for to discrete stokastiske variable X og Y være angivet som tabellen:

y\x	1	2	3
5	0	$\frac{1}{12}$	0
6	$\frac{2}{12}$	0	$\frac{2}{12}$
7	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{2}{12}$
8	0	$\frac{2}{12}$	0

1) Vis at de marginale tæthedsfunktioner for X og Y er givet ved

y	5	6	7	8
$f_Y(y)$	$\frac{1}{12}$	$\frac{4}{12}$	$\frac{5}{12}$	$\frac{2}{12}$

x	1	2	3
$f_X(x)$	$\frac{4}{12}$	$\frac{4}{12}$	$\frac{4}{12}$

Matlab code)

```
XMatrix =[ 1 2 3];
YMatrix = [5;6;7;8];
Matrix = [0 1/12 0; 2/12 0 2/12; 2/12 1/12 2/12; 0 2/12 0];

X = sum(Matrix);
Y = sum(Matrix,2);
format rat
X %Svar
```

```
X =
       1/3            1/3            1/3
```

```
Y % Svar
```

```
Y =
       1/12
       1/3
       5/12
       1/6
```

2) Vis at:

$E[X] = 2, E[Y] = \frac{20}{3}, E[Y \cdot X] = \frac{40}{3}, E[X^2] = \frac{14}{3}, E[Y^2] = \frac{271}{6}$

$$E[X] = \sum x_i * fx(x_i)$$

```
E_x = sum(XMatrix.*X)
```

```
E_x =
       2
```

$$E[y] = \sum y_i * fx(y_i)$$

```
E_Y = sum(YMatrix .* Y)
```

```
E_Y =
     20/3
```

$E[Y * X] = E[X] * E[Y]$

```
EYX=E_x * E_Y
```

```
EYX =
     40/3
```

$$E[X^2] = \sum x_i^2 * fx(x_i)$$

```
EX2 = sum(XMatrix.^2 .* X)
```

```
EX2 =
     14/3
```

$$E[Y^2] = \sum y_i^2 * fx(y_i)$$

```
EY2 = sum(YMatrix.^2 .* Y)
```

```
EY2 =
    271/6
```

3) Hvad er korrelationskoefficenten for X og Y?

Korrelationkoefficenten indikerer hvor meget to tilfældige variabler X og Y er korreleret.

$$\rho = E\left[\frac{X-\bar{X}}{\sigma_X}\cdot\frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}$$

```
VarX = EX2-(E_x.^2);
VarY = EY2 - (E_Y.^2);

p =(EYX - E_x .* E_Y)/(sqrt(VarX)*sqrt(VarY))
```

```
p =
       0
```

4) Angiv om de stokastiske variable X og Y er korrelerede.

- Da korrelationkoefficienten er 0, er X og Y ukorrelerede.

5) Angiv om de stokastiske variable X og Y er uafhængige.

$$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$$

```
inde = kron(Y,X)
```

```
inde =
       1/36           1/36           1/36
       1/9            1/9            1/9
       5/36           5/36           5/36
       1/18           1/18           1/18
```

- Da fx(x)*fy(y) ikke er identisk med fxy(x,y) er X og Y ikke uafhængige.

6) Opskriv den betingede sandsynlighed

$$f_{X|Y}(x|y=6).$$

formel

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = Pr(X=x|Y=y)$$

```
fxy= Matrix(2,:)./Y(2)
```

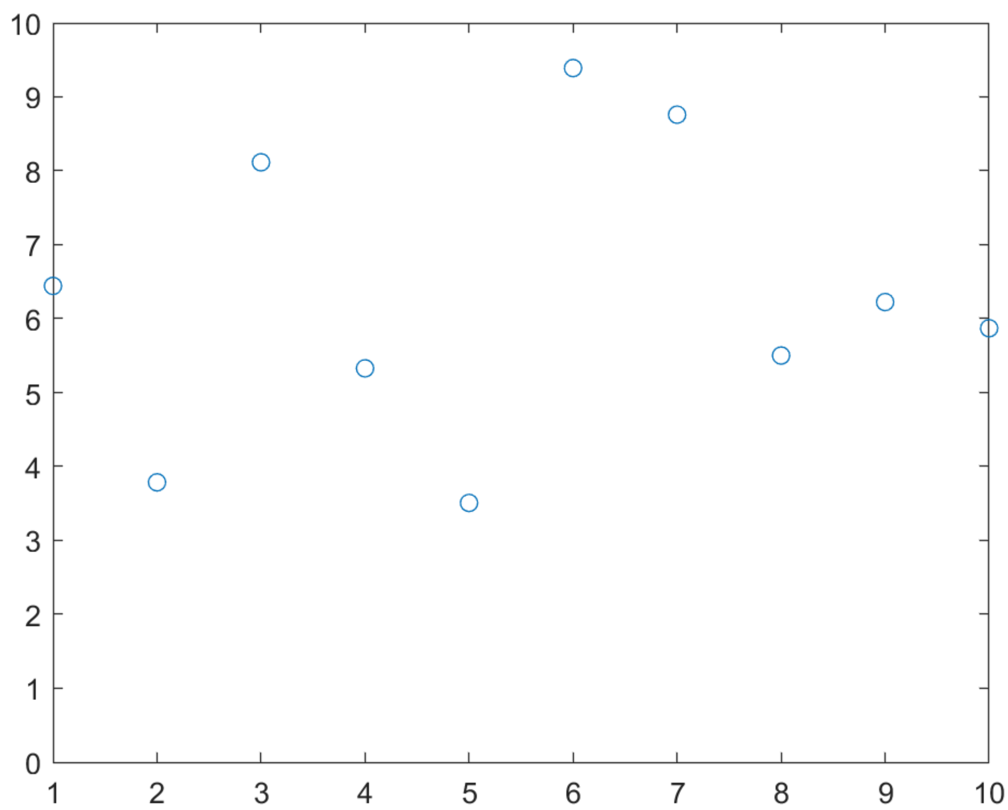
```
fxy =
       1/2             0               1/2
```

Opgave 2

```
clear
clc
format
```

Matlab code for opgave 1

```
for(n = 1:10)
    X(n) = rand * 10;
end
figure(1)
plot(X,'o')
ylim([0 10])
```




```
```

Opgave 2

```
mean = (10+0)/2
```

```
mean = 5
```

```
variance = 1/12*(10-0).^2
```

```
variance = 8.3333
```

Opgave 4 ----

Opgave 1

```
clear
clc

antal = [5562 4357 3471 3078 2309 1285 969 602 238 268];
years = [1901:10:1991];

antalMean = mean(antal)
```

```
antalMean = 2.2139e+03
```

```
yearsMean = mean(years)
```

```
yearsMean = 1946
```

Benytter følgende formel for at finde skæringen og hældningen

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x},$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

```
%Parameter estimat hældning
Beta = sum((antal - antalMean) .* (years - yearsMean)) / sum((years - yearsMean).^2)
```

```
Beta = -59.5000
```

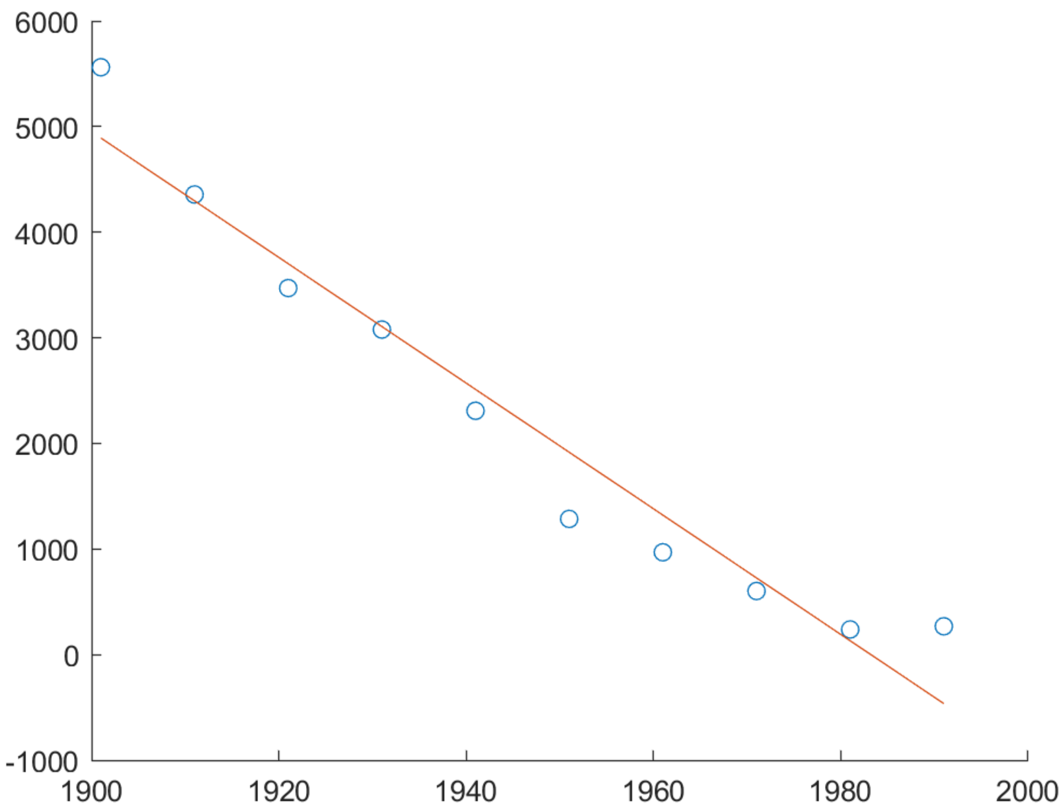
```

% Parameter estimat skæring
Alpha = antalMean - Beta * yearsMean
```

```
Alpha = 1.1800e+05
```

```
linearModel = Alpha + years * Beta;
figure(2)
```

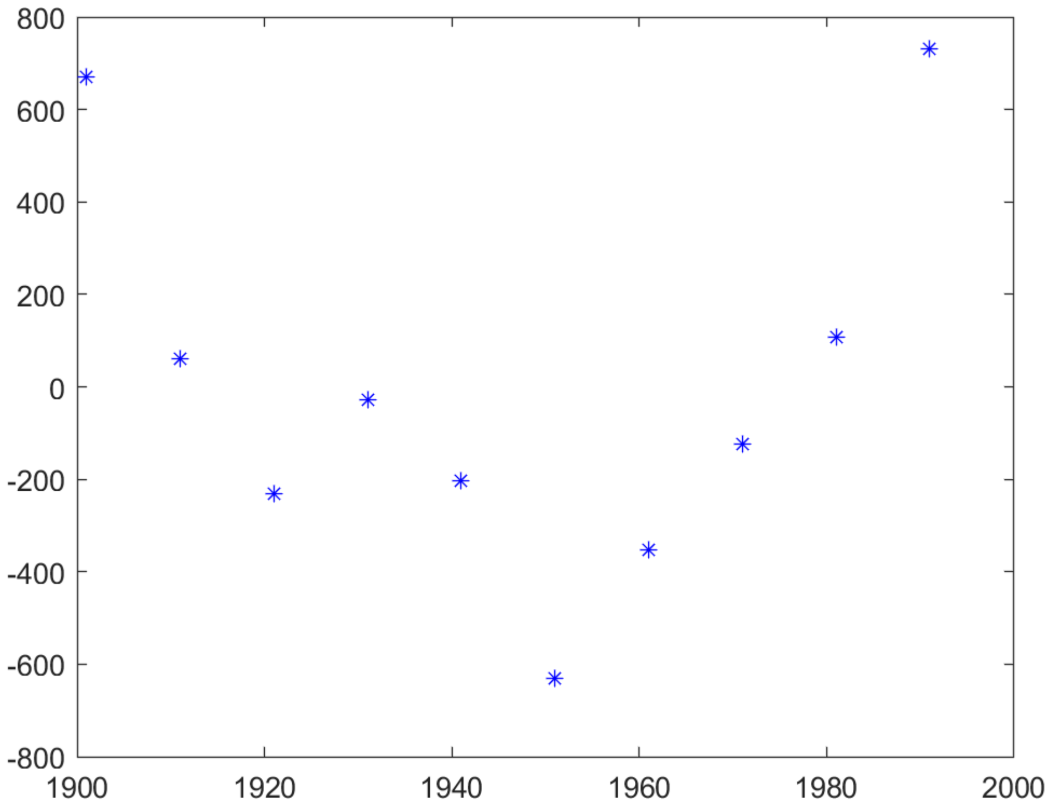
```
hold on
scatter(years, antal)
plot(years,linearModel)
```



2) Lav en residualtegning.

$$\epsilon_i = y_i - (\alpha + \beta x_i)$$

```
Residualtegning = antal - linearModel;
figure(3)
plot(years,Residualtegning,'b*')
```



3) Beregn et 95% konfidensinterval for hældningen.

Først findes standard deviationen

```
n = length(antal)
```

```
n = 10
```

```
sr2 = 1/(n-2)*sum((Residualtegning).^2)
```

```
sr2 = 2.0415e+05
```

$$tinv(0.975, n-2)$$

```
t0 = tinv(0.975,(n-2))
```

```
t0 = 2.2809
```

95% konfidensintervallet kan derefter beregnes med følgende to formler:

$$\hat{\beta}_{-} = \hat{\beta} - t_0 * \sqrt{\frac{s_r^2}{\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2}}$$

$$\hat{\beta}_{+} = \hat{\beta} + t_0 * \sqrt{\frac{s_r^2}{\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2}}$$

```
ConfMinus = Beta - t0 *sqrt(sr2/sum((years - yearsMean).^2))
```

```
ConfMinus = -70.8461
```

```
ConfPositiv = Beta + t0 *sqrt(sr2/sum((years - yearsMean).^2))
```

```
ConfPositiv = -48.1539
```

4) Ud fra svaret i opgave 2 og 3, vil du konkludere at antagelsen om linearitet mellem dødelighed og årstal er rimelig?

Der er umiddelbart for få målepunkter, dog kunne det indiker at der er en lineær tedens, men residual plottet viser at residulaerne systemadisk ligger under 0 mellem årerne 1920-1970, og derudover vil data afvige mere efter 1991, da dødeligeheden ikke kan være negativ.

Linearitet er ikke en god antagelse.

Efterår2015Reeksamen

```
clearvars
close all
```

Opgave 1: Stokastiske Variable

En kontinuert stokastisk variabel X har følgende fordelingsfunktion (cdf):

$$F_X(x) = \begin{cases} k \cdot e^x, & -\infty < x \le 1 \\ 1, & 1 < x \end{cases}$$

1) Vis at tæthedfunktionen (pdf) er givet ved:

$$f_X(x) = \begin{cases} k \cdot e^x, & -\infty < x \le 1 \\ 0, & 1 < x \end{cases}$$

Løst i mathcad prime 4.0

From cdf to pdf: $f_X(x) = \dfrac{dF_X(x)}{dx}$

$$\frac{d}{dx} k \cdot e^x \rightarrow k \cdot e^x \qquad \frac{d}{dx} 1 \rightarrow 0$$

$$f_x(x) := \begin{array}{l} \text{if } -\infty < x \le 1 \\ \quad k \cdot e^x \\ \text{if } 1 < x \\ \quad 0 \end{array}$$

2) For hvilken værdi af k er $f_X(x)$ en gyldig tæthedsfunktion? Begrund svaret.

For at fx(x) er en gyldig tæthedsfunktion gælder følgende:

$$\int_{-\infty}^{\infty} fx(x)dx = 1$$

Samt at alle værdierne af pdf er større eller lig 0

$$kGyldig := \int_{-\infty}^{1} k \cdot e^{x} \, dx + \int_{1}^{\infty} 0 \, dx = 1 \xrightarrow{solve, k} e^{-1}$$

3) Brug $F_X(x)$ til at beregne sandsynlighederne $\Pr(x < 0{,}4)$ og $\Pr(0{,}1 \leq x < 0{,}4)$. Antag at $k = \frac{1}{e}$.

$$Pr1 := F_x(0.4) \rightarrow 1.4918246976412703178 \cdot k = 0.549$$

$$Pr2 := F_x(0.1) \rightarrow 1.1051709180756476248 \cdot k = 0.407$$

$$Pr_{svar} := Pr1 - Pr2 = 0.142$$

4) Bestem forventningsværdien og variansen af X udfra $f_X(x)$. Antag at $k = \frac{1}{e}$.

Forventningsværdien givet ved

$$E\ |X| = \int x \cdot f(x)\,\mathrm{d}x$$

$$Forventningsværdi$$

$$forventningsværdi := \int_{-\infty}^{1} x \cdot \left(k \cdot e^{x}\right)\,\mathrm{d}x \rightarrow 0$$

Variance er givet ved

$$Var(x) = E\ \left|x^2\right| - E\ |X|^2$$

$$forventningsIAnden := \int_{-\infty}^{1} x^2 \cdot \left(k \cdot e^{x}\right)\,\mathrm{d}x$$

$$varianse := forventningsIAnden - forventningsværdi^2 = 1$$

Opgave 2: Stokastiske Processer

En diskret stokastisk process er givet ved:

$$X(n) = w(n) + 4$$

Hvor hver sample n af w er i.i.d Gaussisk fordelte stokastiske variable $w(n) \sim N(0{,}1)$.

1) Skitser 10 samples ($n = 1{,}2, \ldots, 10$) af en realisation af processen $X(n)$.

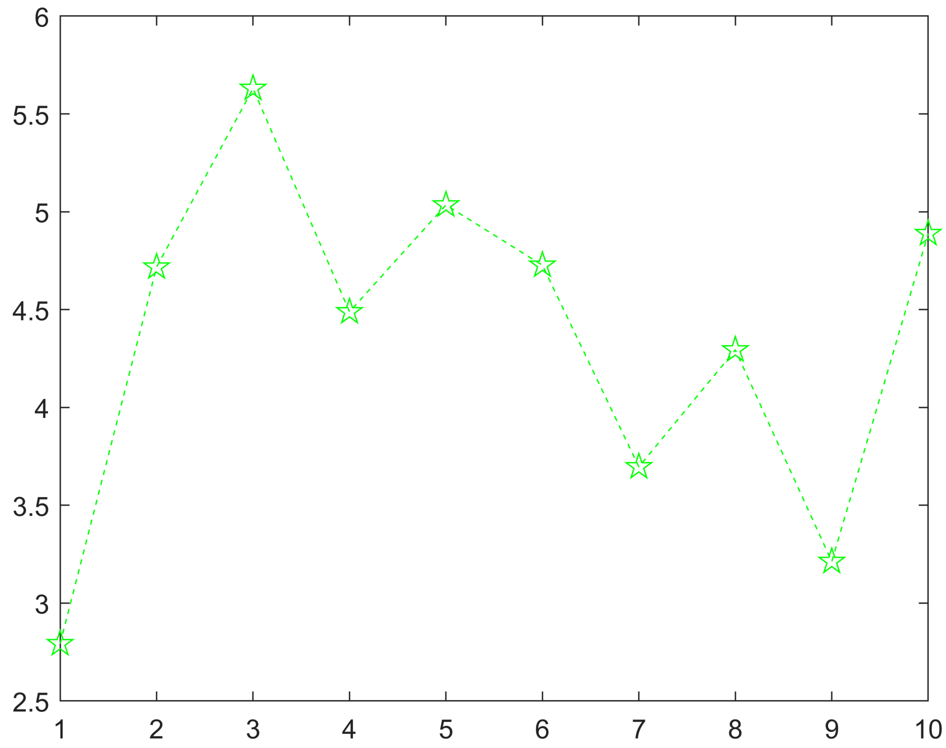
```
clear all
clc
n = 1:1:10
```

```
n = 1×10
     1     2     3     4     5     6     7     8     9    10
```

```

for(n = 1:10)
    X(n) = randn+4;
```

```
end
figure(1)
plot(X,'--gp','MarkerSize',10)
```



2) Bestem ensemble middelværdien og ensemble variansen for processen $X(n)$.

Da ensemble middelværdi ved w(n)~N(0,1) er 0 pga normalfordelingsfunktionen, og +4 ved den givne stokastiske process

Så gives dette

```
EnsembleMean = 0 + 4
```

```
EnsembleMean = 4
```

Variance er givet ved

$$\sigma_x^2 = E|x^2| - E|x|^2$$

ensemble variansen ved w(n)~N(0,1) er 1 pga normalfordelingsfunktionen

: $E[w(n)^2] + E[4^2] - 4^2$

```
ensembleMean = 1^2+4^2-EnsembleMean^2
```

```
ensembleMean = 1
```

3) Angiv om processen er WSS (stationær i den brede forstand) og om den er ergodisk. Begrund dine svar.

processen er WSS, da middelværdi og varians ikke er tidsafhængige.
processen er ergodisk, da middelværdi og varians kan bestemmes udfra én realization.

Opgave 3: Sandsynlighedsregning

Du spiller kort i et kasino. Der er et sæt af 52 kort i et spil. Du trækker syv kort.

1) Hvis hændelse A er at hjerter konge er blandt de syv kort. Hvad er sandsynligheden for hændelse A?

```
Pr_a =  1/52 + 1/52 + 1/52 + 1/52 + 1/52 + 1/52 + 1/52
```

```
Pr_a = 0.1346
```

2) Hvis hændelse B er at spar es er blandt de syv kort. Hvad er den simultane sandsynlighed for hændelserne A og B?

```
Pr_b =  Pr_a
```

```
Pr_b = 0.1346
```

```
PrBgivetA = 6 * 1/51
```

```
PrBgivetA = 0.1176
```

```
PrA_B = PrBgivetA * Pr_b
```

```
PrA_B = 0.0158
```

3) Er hændelserne A og B uafhængige? Begrund dit svar.

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

```
eq(Pr_a*Pr_b,PrA_B)
```

```
ans = logical
   0
```

Da disse er ikke ens, så kan vi bekræfte at de ikke er uafhænige.

4) Hvor mange forskellige kombinationer af 7 kort kan der trækkes fra et spil kort af 52 kort?

The number of combinations is:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

```
factorial(52)/(factorial(7)*factorial(52-7))
```

```
ans = 133784560
```

Opgave 4: Statistik

Den gennemsnitlige alder for 1. gangs viede mænd i Danmark er angivet ved følgende tabel:

Alder:	25,2	26,5	27,9	29,2	30,2	31,7	32,8	34,0	34,3
År:	1971	1976	1981	1986	1991	1996	2001	2006	2011

Alder er den gennemsnitlige af mænd, der bliver gift for første gang i årstalet angivet ved År[1]:

1) Plot data fra tabellen. Anvend lineær regression til at bestemme en model for data, angiv hvorledes modellens parametre er beregnet (angiv desuden formlerne, der er brugt ved beregningen). Indtegn desuden den rette linie på plottet.

Alder = [25.2 26.5 27.9 29.2 30.2 31.7 32.8 34.0 34.3]

Years = 1971:5:2011

```
Alder = [25.2 26.5 27.9 29.2 30.2 31.7 32.8 34.0 34.3];
Years = 1971:5:2011;
alderMean = mean(Alder)
```

```
alderMean = 30.2000
```

```
yearsMean = mean(Years)
```

```
yearsMean = 1991
```

```

```

Benytter følgende formel for at finde skæringen og hældningen

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x},$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

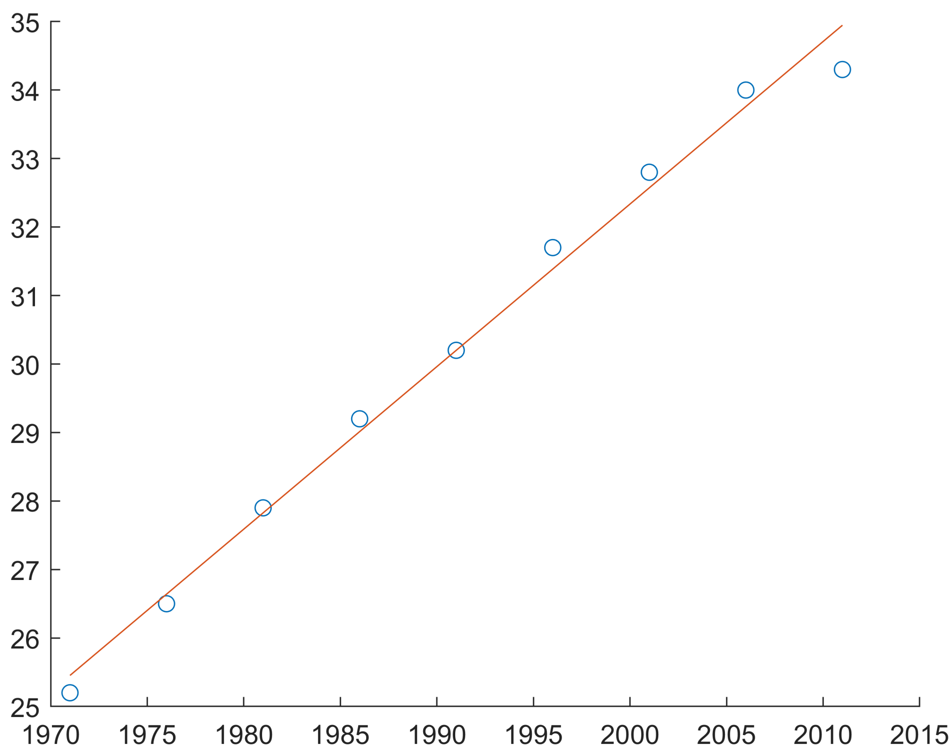
```
%Parameter estimat hældning
Beta = sum((Alder - alderMean) .* (Years - yearsMean)) / sum((Years - yearsMean).^2)
```

```
Beta = 0.2373
```

```
% Parameter estimat skæring
Alpha = alderMean - Beta * yearsMean
```

```
Alpha = -442.3307
```

```
linearModel = Alpha + Years * Beta;
figure(2)
hold on
scatter(Years, Alder)
plot(Years,linearModel)
```

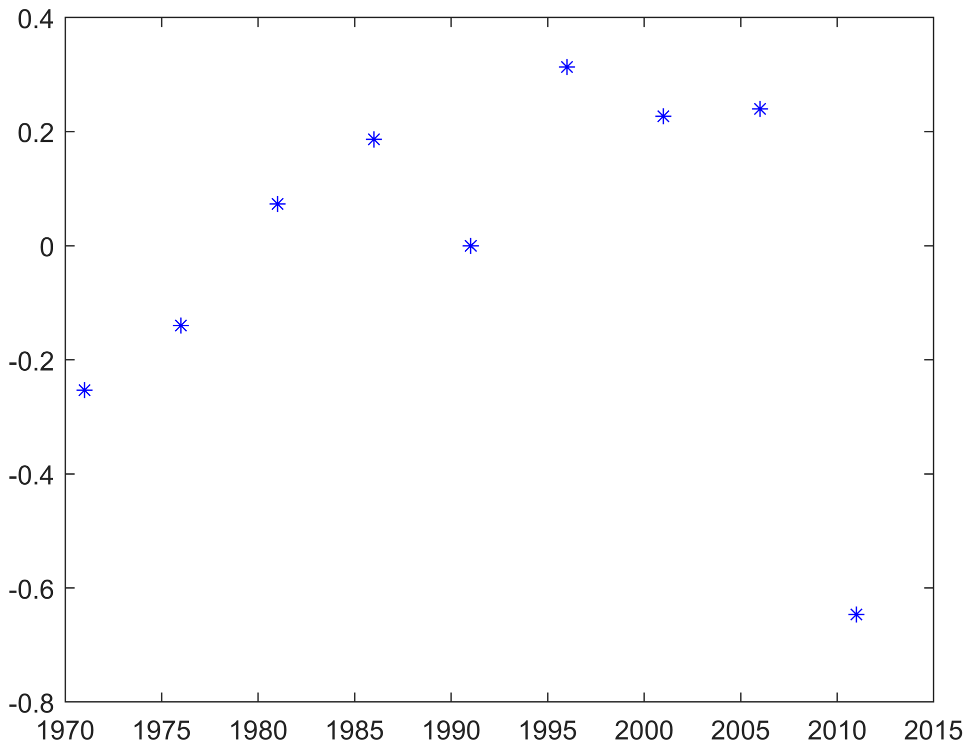


2) Lav en residualtegning på en graf. Angiv desuden hvordan residualerne på grafen beregnes (angiv desuden formlerne, der er brugt ved beregningen).

$$\epsilon_i = y_i - (\alpha + \beta x_i)$$

```
Residualtegning = Alder - linearModel;
figure(3)
```

```
plot(Years,Residualtegning,'b*')
```



3) Antag at det samlede antal vielser med en mand der bliver gift for første gang er 20.000 pr. År. Antag desuden at data er normalfordelt. Hvis du skal sammenligne de to middelværdier for aldrene på 1.gangs viede mænd i henholdsvis år 1971 og 2011, hvilken statistisk test vil du benytte? Begrund dit svar.

Da varians. er ukendt og data er Gaussisk og ikke parret, bruges en t-test for sammenligning af middelværdier.

Efterår2016

```
clearvars
clc
close all
```

Opgave 1: Stokastiske Variable

En diskret stokastisk variabel X har følgende tæthedssfunktion (pmf):

x	-1	1	7
$f_X(x)$	k	$\frac{3}{4}$	$\frac{1}{8}$

1) For hvilken værdi af k er $f_X(x)$ en gyldig tæthedsfunktion? Begrund svaret.

```
clearvars
```

For at det er en gyldig tæthedsfunktion er følgende givet:

$$\sum_x fx(x) = 1$$

```
syms k
x=[-1 1 7]
```

```
x = 1×3
    -1     1     7
```

```
fx = [ k 3/4 1/8]
```

fx =

$$\left(k \;\; \frac{3}{4} \;\; \frac{1}{8}\right)$$

```
k = solve(sum(fx)== 1 , k)
```

k =

$$\frac{1}{8}$$

2) Antag at $k = \frac{1}{8}$, find fordelings funktionen (cdf) $F_X(x)$ for X. Skitsér $F_X(x)$.

$$Fx(x) = \int_{nedre\ værdi}^{x} x\,dx + tidl.\ led$$

```
syms x
Fx(x) = piecewise( x < -1, 0, ...
   x           -1 <= x < 1, k, ...
              1 <= x < 7, k+3/4, ...
              7 <= x, k + 3/4 + 1/8)
```

Fx(x) =

$$\begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{8} & \text{if } x < 1 \\ \frac{7}{8} & \text{if } x \in [1,7) \\ 1 & \text{if } 7 \le x \end{cases}$$

```
fplot(Fx, [-5, 10],'-r','LineWidth',2)
```

3) Brug $f_X(x)$ til at finde forventningsværdien $E[X]$ og standard afvigelsen σ_X for X. Antag at $k = \frac{1}{8}$.

Ved forventningsværien gælder følgende:

$$\boldsymbol{E}|\boldsymbol{x}| = \sum_{x} x * fx(x)$$

```
fx = [ 1/ 8 3/4 1/8]
```

```
fx = 1×3
    0.1250    0.7500    0.1250
```

```
x=[-1 1 7]
```

```
x = 1×3
    -1     1     7
```

```
Ex= sum(x.*fx)
```

```
Ex = 1.5000
```

Variance er givet ved

$$Var(x) = E\left[x^2\right] - E\left[X\right]^2$$

```
Ex2= sum(x.^2.*fx)
```

```
Ex2 = 7
```

```
var = Ex2 - Ex^2
```

```
var = 4.7500
```

4) Hvis en funktion er defineret som $g(X = x) = 3 \cdot x^2$. Find forventningsværdien $E[g(X = x)]$. Antag at $k = \frac{1}{8}$.

```
E_G_3_X2= 3*Ex2
```

```
E_G_3_X2 = 21
```

5) Angiv hvilke værdier X kan antage.

X kan antage -1, 1 og 7

Opgave 2: Stokastiske Processer

En kontinuer stokastisk process er givet ved:

$$X(t) = w(t)$$

Hvor $w(t)$ er i.i.d uniformt fordelt efter $w(t) \sim U(-2, -1)$.

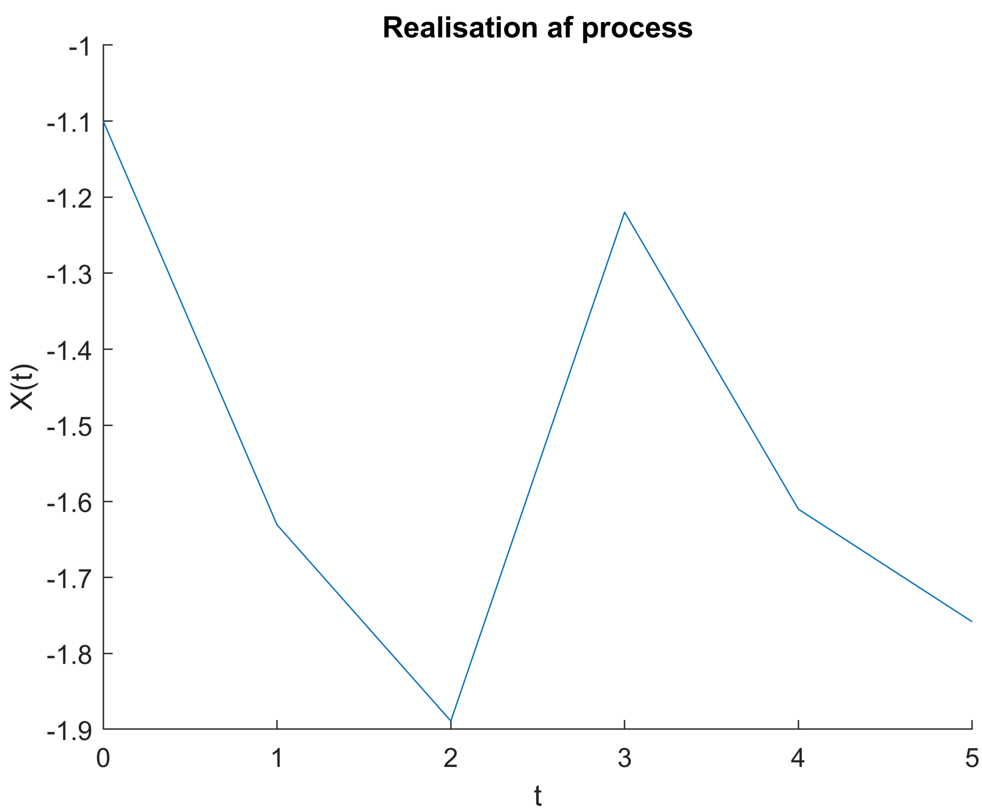
1) Den stokastiske process $X(t)$ er samplet hvert sekundt, skitsér 6 samples fra 0 - 5 s af én realisation. Angiv hvorledes realisationen er fremkommet, brug en tilfældighedsgenerator, f.eks. rand() i matlab.

```
clearvars
```

```
close all
t = 0:1:5;
samples = length(t);
a = -2;
b = -1
```

```
b = -1
```

```
w = rand(1,samples)-2;
figure(1)
hold on
xlabel("t"); ylabel("X(t)"); title("Realisation af process");
plot(t, w)
```



2) Bestem ensemble middelværdien og ensemble variansen for processen $X(t)$.

Mean value: $\mu = \frac{a+b}{2}$

Variance: $\sigma^2 = \frac{1}{12}(b-a)^2$

```
a= -2;
b = -1;
format rat
```

```
middel = (a+b)/2
```

```
middel =
      -3/2
```

```
variance2=1/12*(b-a)^2
```

```
variance2 =
       1/12
```

3) Opskriv formlen til at bestemme den tidslige middelværdi for processen $X(t)$.

$$\hat{\mu}_{X_i} = \langle X_i \rangle_T = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt$$

4) Angiv om processen $X(t)$ er WSS (stationær i den brede forstand) og om den er ergodisk. Begrund dine svar.

Processen er WSS da middelværdien og variansen ikke er tidsafhængig

Processen er ergodisk da tæthedsfunktionen kan findes ud fra en realistation når t -> uendelig

Opgave 3: Sandsynlighedsregning

En HIV test baseret på spyt er positiv i 92% af tilfældene, givet at man er HIV smittet. Den samme test er negativ i 98% af tilfældene, givet at man er ikke er HIV smittet. Af hele befolkningen er 0,1% smittet med HIV.

1) Hvad er sandsynligheden for at en person fra befolkningen både er HIV smittet og har en positiv test?

Hændelse A : Hiv smittet

Hændelse B = test positiv

```
format
Pr_H_T = 0.92;
```

```
Pr_NOT_H_NOT_T = 0.98;
Pr_H = 0.001;

Pr_hiv_test_positiv = Pr_H_T*Pr_H
```

```
Pr_hiv_test_positiv = 9.2000e-04
```

2) Hvad er den totale sandsynlighed for at en person fra befolkningen har en positiv test?

Positiv test = 92

Raske som får en positiv er 1-0.98 = 0.02

```
Pr_H_NOT_T=1-Pr_NOT_H_NOT_T;
Pr_NOT_H = 1-Pr_H;

Pr_T = Pr_NOT_H*Pr_H_NOT_T+Pr_hiv_test_positiv
```

```
Pr_T = 0.0209
```

3) Hvis en person fra befolkningen har en positiv test, hvad er sandsynligheden for, at han er HIV smittet?

```
Pr_hiv_test_positiv/Pr_T
```

```
ans = 0.0440
```

4) Er begivenhederne ”At have en positv test” og ”være HIV smittet” uafhængige? begrund dit svar.

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

```
Pr_T*Pr_H
```

```
ans = 2.0900e-05
```

```
Pr_hiv_test_positiv
```

```
Pr_hiv_test_positiv = 9.2000e-04
```

Da overstående formel ikke er gældende, så kan vi bekræfte at de ikke er uafhængige.

Opgave 4: Statistik

Antal patienter døde af AIDS i DK mellem 1985 - 1994 er angivet ved tabellen. Antal døde er angivet ved "Antal" og årstalet er angivet ved "År"[1].

Antal:	28	46	44	63	104	148	172	187	223	236
År:	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994

1) Hvad er den empiriske middelværdi og den empiriske varians for antallet af døde AIDS patienter?

```
antal = [28 46 44 63 104 148 172 187 223 236];
years = 1985:1:1994;
emperiskeMiddel = mean(antal)
```

```
emperiskeMiddel = 125.1000
```

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

```
empiriskeVarianse2= sum((antal-emperiskeMiddel).^2)/(length(antal)-1)
```

```
empiriskeVarianse2 = 6.1114e+03
```

2) Plot data fra tabellen. Anvend lineær regression til at bestemme en model for data, angiv hvorledes modellens parametre er beregnet (skæringen med y-aksen og hældningen af den lineære model). Indtegn desuden den rette linie på plottet.

```
yearsMean = mean(years)
```

```
yearsMean = 1.9895e+03
```

Benytter følgende formel for at finde skæringen og hældningen

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x},$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

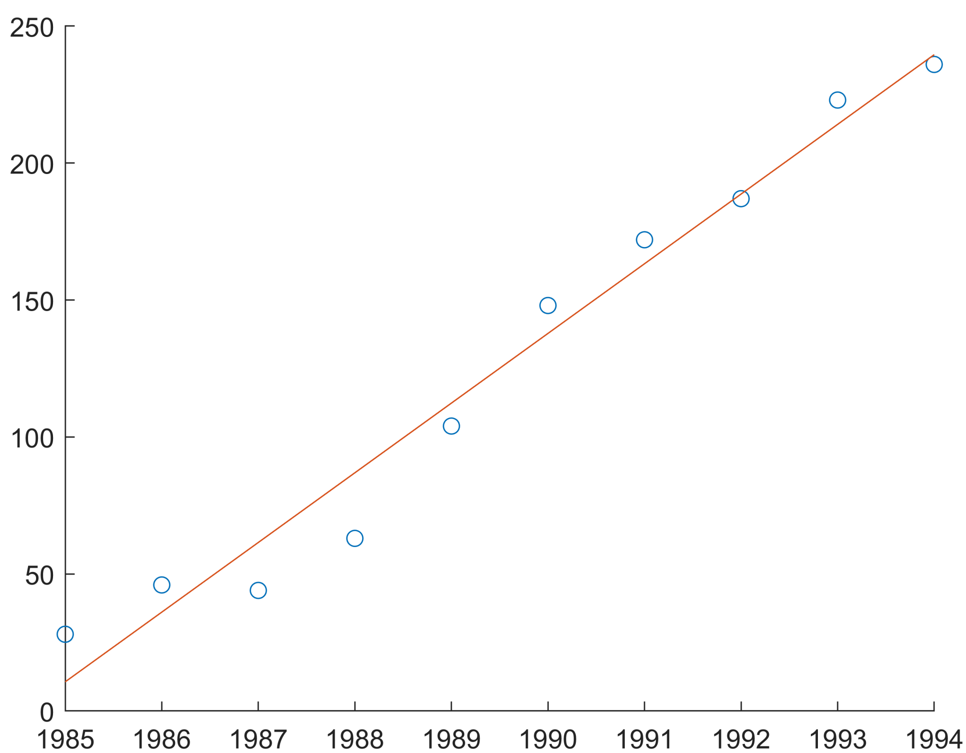
```
%Parameter estimat hældning
Beta = sum((antal - emperiskeMiddel) .* (years - yearsMean)) / sum((years - yearsMean).^2)
```

```
Beta = 25.4364
```

```
% Parameter estimat skæring
Alpha = emperiskeMiddel - Beta * yearsMean
```

```
Alpha = -5.0481e+04
```

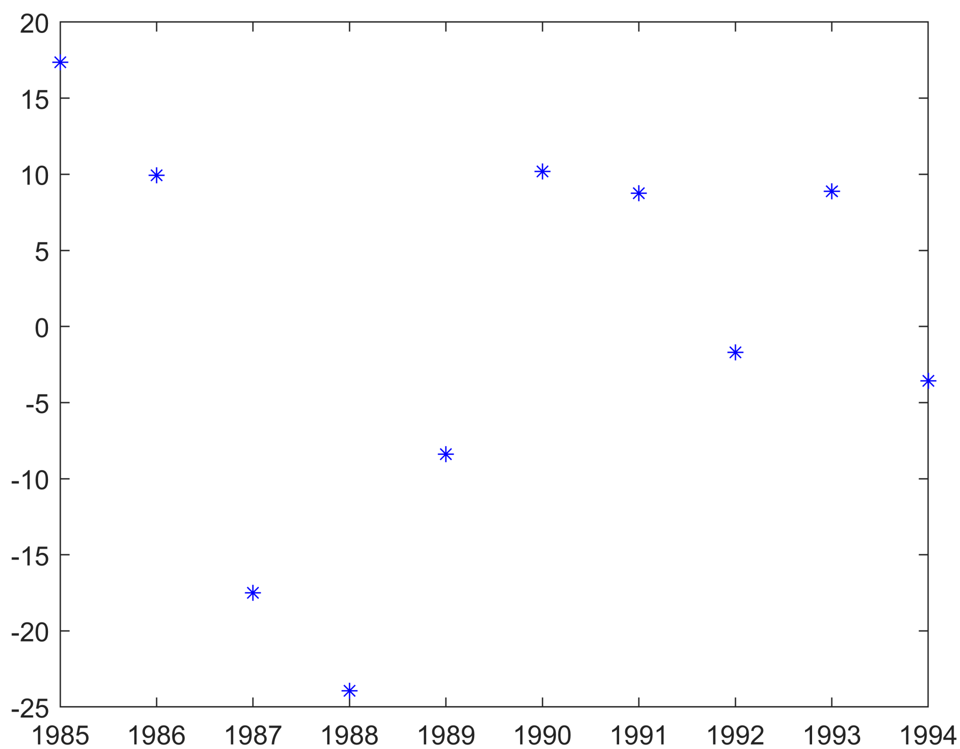
```
linearModel = Alpha + years * Beta;
figure(2)
hold on
scatter(years, antal)
plot(years,linearModel)
```



3) Lav en residualtegning for modellen fra 2) på en graf. Angiv desuden hvordan residualerne på grafen beregnes.

$$\epsilon_i = y_i - (\alpha + \beta x_i)$$

```
Residualtegning = antal - linearModel;
figure(3)
plot(years,Residualtegning,'b*')
```



4) Beregn et 95% konfidensinterval for hældningen.

Først findes standard deviationen

$tinv(0.975, n-2)$

```
n = length(antal)
```

```
n = 10
```

```
sr2 = 1/(n-2)*sum((Residualtegning).^2)
```

```
sr2 = 203.0864
```

```
t0 = tinv(0.975,(n-2))
```

```
t0 = 2.3060
```

95% konfidensintervallet kan derefter beregnes med følgende to formler:

$$\hat{\beta}_{-} = \hat{\beta} - t_0 * \sqrt{\frac{s_r^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}}$$

$$\hat{\beta}_{+} = \hat{\beta} + t_0 * \sqrt{\frac{s_r^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}}$$

```
ConfMinus = Beta - t0 *sqrt(sr2/sum((years - yearsMean).^2))
```

```
ConfMinus = 21.8183
```

```
ConfPositiv = Beta + t0 *sqrt(sr2/sum((years - yearsMean).^2))
```

```
ConfPositiv = 29.0544
```

5) Udfra svaret i opgave 3) og 4), vil du konkludere at antagelsen om linearitet mellem antal døde og årstal er rimelig? Begrund dit svar.

Udfra residualplottet er residualerne fordelt pænt omkring 0. Der er altså ikke et tydeligt mønster. Desuden ligger konfidensintervallet mellem 21.8183 og 29,0544 og altså tæt. Antagelsen om linearitet virker derfor rimelig.

6) Er der nogle årstal, hvor den lineære model ikke kan bruges?

Årstal hvor antallet er negativt kan ikke bruges.

```
clearvars
clc
close all
```

Opgave 1: Stokastiske Variable

En kontinuert stokastisk variabel X har følgende tæthedsfunktion (pdf):

$$f_X(x) = \begin{cases} A \cdot x + B & -2 \leq x \leq 3 \\ 0 & ellers \end{cases}$$

1) En skitse af $f_X(x)$ er som på figur 1, hvor det bemærkes at $f_X(3) = 0$.
For hvilken værdi af $f_X(-2) = k$, er $f_X(x)$ en gyldig tæthedsfunktion?
Begrund dit svar.

```
clearvars
```

For at fx(x) er en gyldig tæthedsfunktion gælder følgende:

$$\int_{-2}^{3} fx(x)dx = \quad 1$$

Areal skal under kurven være 1, derved bruger vi formlen fra en retvinklet trekant.

$$Areal = 1 = \frac{1}{2} * k * 5 \rightarrow \quad k = \frac{2}{5} \quad = \quad 0.4$$

2) Vis at fordelingsfunktionen (cdf) $F_X(x)$ for X er givet ved:

$$F_X(x) = \begin{cases} 0 & x < -2 \\ \frac{A}{2} \cdot x^2 + B \cdot x + C & -2 \leq x \leq 3 \\ 1 & 3 < x \end{cases}$$

Antag at $A = -\frac{2}{25}$ og $B = \frac{6}{25}$ og $C = \frac{16}{25}$.

$$Fx(x) = \int_{nedre\ værdi}^{x} x\, dx + tidl.\ led$$

- From pdf to cdf: $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = Pr(X \leq x)$
- From cdf to pdf: $f_X(x) = \frac{dF_X(x)}{dx}$

$$Fx(x) := \frac{a}{2} \cdot x^2 + b \cdot x + c$$

$$\frac{d}{dx} Fx(x) \rightarrow b + a \cdot x$$

3) Brug $f_X(x)$ til at finde forventningsværdien $E[X]$ og variansen ${\sigma_X}^2$ for X.

Antag at $A = -\frac{2}{25}$ og $B = \frac{6}{25}$.

Ved
forventningsværien gælder følgende:

$$\boldsymbol{E|x| = \int_{-\infty}^{\infty} x * fx(x)dx}$$

```
clearvars
syms x
A=-2/25; B = 6 /25;
fx(x) = piecewise( x < -2, 0,...
                   -2 <= x <= 3, A*x+B,...
                   x > 3 ,0)
```

fx(x) =

$$\begin{cases} 0 & \text{if } x < -2 \\ \frac{6}{25} - \frac{2x}{25} & \text{if } x \in [-2, 3] \\ 0 & \text{if } 3 < x \end{cases}$$

```
ex =  int(x*fx(x), -Inf, Inf)
```

ex =

$$-\frac{1}{3}$$

```
ex2 = int(x.^2*fx(x), -Inf, Inf)
```

ex2 =

$$\frac{3}{2}$$

$$\boldsymbol{var(x) = E\left|x^2\right| - E|x|^2}$$

```
variance =  ex2-ex^2
```

variance =

$$\frac{25}{18}$$

4) Hvad er sandsynligheden $\Pr(X < 0)$? Begrund dit svar.

Antag at $A = -\frac{2}{25}$ og $B = \frac{6}{25}$ og $C = \frac{16}{25}$.

Følgende gælder:

$$Pr(x < 0) = Fx(0), \qquad Pr(x < 0{,}4) = \frac{A}{2} * x^2 + B * x + C$$

```
C = 16/25
```

```
C = 0.6400
```

```
Fx(x) = piecewise( x < -2, 0,...
                   -2 <= x <= 3, A/2*x^2+B*x+C,...
                   x > 3 ,1)
```

```
Fx(x) =
```

$$\begin{cases} 0 & \text{if } x < -2 \\ -\frac{x^2}{25} + \frac{6\,x}{25} + \frac{16}{25} & \text{if } x \in [-2, 3] \\ 1 & \text{if } 3 < x \end{cases}$$

```
Pr1 = Fx(0)
```

```
Pr1 =
```

$$\frac{16}{25}$$

Opgave 2: Stokastiske Processer

En diskret stokastisk process er givet for den n´te sample ved:

$$X(n) = W(n) + 0{,}7$$

hvor $W(n)$ er i.i.d. fordelte efter:

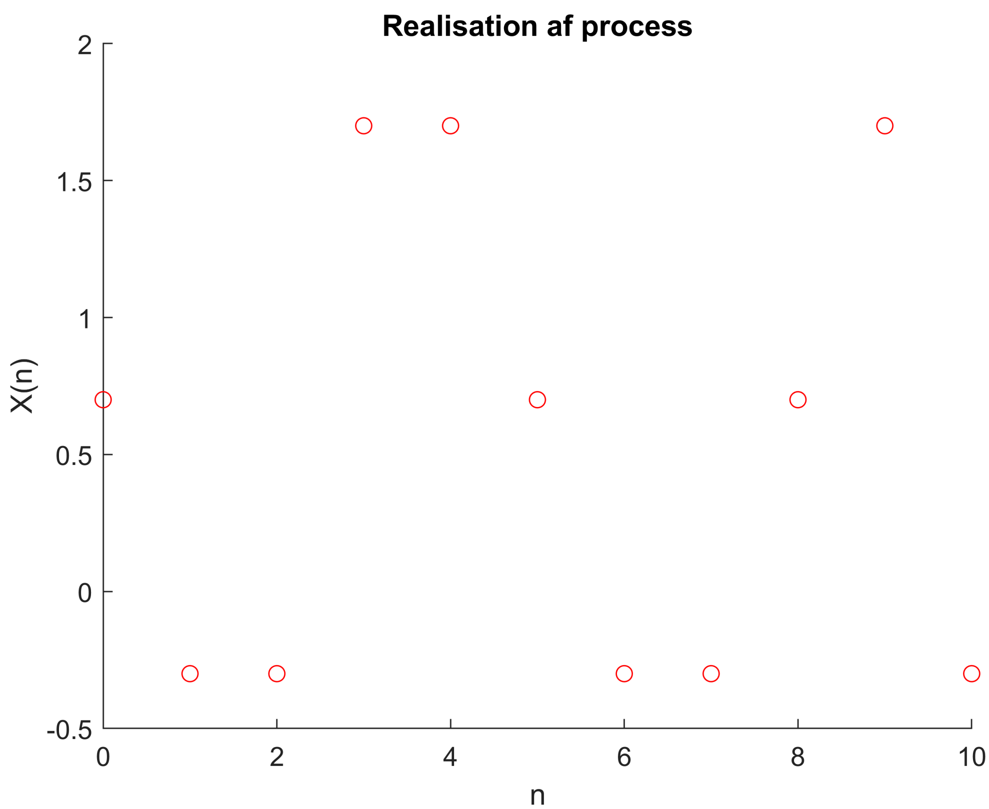
$w(n)$	-1	0	1
$f_{W(n)}(w(n))$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

1) Skitsér 11 samples fra 0 – 10 af én realisation af $X(n)$. Angiv hvorledes realisationen er fremkommet, brug en tilfældighedsgenerator, f.eks. unidrnd() i matlab.

$\mathcal{N}(\mu,\sigma^2)$
Mean value: μ
Variance: σ^2

```
x = 0:1:10;
samples = length(x);
```

```
w = unidrnd(3, 1, samples) - 2 + 0.7;
figure(1);
hold on
xlabel("n"); ylabel("X(n)"); title("Realisation af process");
scatter(x, w, 'or')
```



2) Bestem ensemble middelværdien og ensemble variansen for processen $X(n)$.

$$\boldsymbol{E}|\boldsymbol{x}| = \sum_{x} w * fw(w) + process\ værdi$$

```
x = [-1 0 1]
```

```
x = 1×3
    -1     0     1
```

```
fx= [1/3 1/3 1/3]
```

```
fx = 1×3
    0.3333    0.3333    0.3333
```

```
ex = sum(x.*fx)+0.7
```

```
ex = 0.7000
```

```
ex2= sum(x.^2.*fx)
```

```
ex2 = 0.6667
```

```
%%%% OBS %%%% Variance giver ikke rigtige svar
```

Ved variansen gælder følgende:

$$\sigma_x^2 = E|x^2| - E|x|^2$$

$$\sigma_x^2 = E|x^2| - E|x|^2$$

$$\boldsymbol{E|x^2|} = \sum_x w^2 * fw(w) + 0{,}7 \rightarrow (-1)^2 * \frac{1}{3} + 0^2 * \frac{1}{3} + 1^2 * \frac{1}{3} = \frac{2}{3}$$

$$\sigma_x^2 = E|x^2| - E|x|^2 = \frac{2}{3}$$

```
variance2 = ex2 - ex^2
```

```
variance2 = 0.1767
```

3) Opstil formlen til at bestemme autokorrelationsfunktionen for $X(n)$.

$$\mathrm{Rxx}(\tau) = \sum x(n)x(n-\tau) * \mathrm{fx}(x)$$

hvor $f(x) = \mathrm{fw}(w + 0.7)$

4) Angiv om processen $X(n)$ er WSS (stationær i den brede forstand), og om den er ergodisk. Begrund dine svar.

Da processen er uafhængig af tiden, og at middelværdien og variansen ikke ændrer sig over tid, så kan vi konstanter at processen er stationær.

Processen er egodisk da pmf kan bestemmes udfra en realisation

Opgave 3: Sandsynlighedsregning

I et tognet angives forsinkelser til at kunne skyldes blade på skinnerne, signalfejl eller personalemangel. Hændelserne er uafhængige og er ikke disjunkte.
Hvis der opstår en forsinkelse, vil der være blade på skinnerne $\frac{1}{4}$ af gangene, der vil være signalfejl $\frac{1}{2}$ af gangene, og der vil være personalemangel $\frac{1}{4}$ af gangene.

Haendelse A : Blade på skinnerne

Haendelse B: Signal fejl

Haendelse C: Personmangel

1) Skitsér Venn diagrammet (hændelsesdiagram for udfaldsrummet) for de tre begivenheder, når blade på skinnerne er hændelse A, signalfejl er hændelse B, og personalemangel er hændelse C.

2) Find sandsynligheden $\Pr(A \cap B)$. Dvs. sandsynligheden for, at der både er blade på skinnerne og signalfejl.

```
Pr_A = 1/4;
Pr_B = 1/2;
Pr_C = 1/4;

Pr_A_B = Pr_A*Pr_B
```

```
Pr_A_B = 0.1250
```

3) Find sandsynligheden $\Pr(A \cup B)$. Dvs. sandsynligheden for, at der enten er blade på skinnerne eller signalfejl.

Union: $$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

```
Pr_AUB = Pr_A + Pr_B - Pr_A_B
```

```
Pr_AUB = 0.6250
```

4) Find sandsynligheden $\Pr(A \cup B \cup C)$. Dvs. sandsynligheden for, at en forsinkelse på toget skyldes blade på skinnerne, signalfejl eller personalemangel.

Da A, B og C er uafhængige så tages der Pr(A U B U C)

```
Pr_A + Pr_B + Pr_C - Pr_A_B - (Pr_B*Pr_C) - (Pr_A*Pr_C) + (Pr_A*Pr_B*Pr_C)
```

```
ans = 0.7188
```

Opgave 4: Statistik

I et studie af en anerkendt metode til vægttab undersøges 10 patienter før behandling, og ét år efter behandling.

Patient nr.	Vægt før (kg)	Vægt efter (kg)
1	140	130
2	138	121
3	110	127
4	154	101
5	125	92
6	169	170
7	142	143
8	162	170
9	131	134
10	122	85

1) Opstil en NULL og Alternativ hypotese for at bestemme, om behandlingen har ændret patienternes vægt.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

$H_1 = \mu 1 \neq \mu 2$

2) Bør testen, der udføres, være parret eller uparret? Begrund dit svar.

Da det er samme subjekt for før og efter, så bør det være en parret.

3) Estimér middelforskellen af patienternes vægt før og efter behandling.

```
patient = 1:1:10;
foer = [ 140 138 110 154 125 169 142 162 131 122];
efter = [130 121 127 101 92 170 143 170 134 85];

middelforskel = sum(efter-foer)/length(patient)
```

```
middelforskel = -12
```

4) Estimér variansen for forskellen af patienternes vægt før og efter behandling.

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

```
s2 = 1/(length(patient)-1)*(sum((efter-foer-middelforskel).^2))
```

```
s2 = 508.8889
```

5) Anvend en parret t-test til hypotesetest af din hypotese. Kan NULL hypotesen afvises med et signifikansniveau på 0,05? Begrund dit svar.

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

```
t = middelforskel/sqrt(s2/length(patient))
```

```
t = -1.6822
```

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

```
pval = 2*(1-tcdf(abs(t),(length(patient)-1)))
```

```
pval = 0.1268
```

Da pval er større end 0.05, så fejler vi med at afvise hypotese 0.

6) Opstil og find 95% konfidensintervallet for vægtforskellen før og efter behandlingen. Angiv hvilken formel, der er brugt.

Lower bound:

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1}}$$

Upper bound:

$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1}}$$

```
t0=tinv(0.975,(length(patient)-1))
```

```
t0 = 2.2622
```

```
koefMinus = middelforskel-t0*sqrt(s2)*sqrt((1/length(patient)))
```

```
koefMinus = -28.1374
```

```
koefPositiv = middelforskel + t0*sqrt(s2)*sqrt((1/length(patient)))
```

```
koefPositiv = 4.1374
```

```
clearvars
clc
close all
```

Opgave 1: Stokastiske Variable

En kontinuert stokastisk variabel X har følgende fordelingsfunktion (cdf):

$$F_X(x) = \begin{cases} 0, & 2 \geq x \\ k \cdot x - \frac{2}{3}, & 2 < x \leq 5 \\ 1, & 5 < x \end{cases}$$

1) Vis at tæthedssfunktionen (pdf) er givet ved:

$$f_X(x) = \begin{cases} 0, & 2 \geq x \\ k, & 2 < x \leq 5 \\ 0, & 5 < x \end{cases}$$

From cdf to pdf: $f_X(x) = \frac{dF_X(x)}{dx}$

```
syms x k
Fx(x) = piecewise(2 >= x, 0, 2 < x <= 5, k * x -2/3, 5 < x, 1)
```

Fx(x) =

$$\begin{cases} 0 & \text{if } x \leq 2 \\ k\,x - \frac{2}{3} & \text{if } x \in (2, 5] \\ 1 & \text{if } 5 < x \end{cases}$$

```
fx(x) = diff(Fx(x))
```

fx(x) =

$$\begin{cases} 0 & \text{if } x < 2 \\ k & \text{if } x \in (2, 5) \\ 0 & \text{if } 5 < x \end{cases}$$

2) For hvilken værdi af k er $f_X(x)$ en gyldig tæthedsfunktion? Begrund svaret.

For at fx(x) er en gyldig tæthedsfunktion gælder følgende:

$$\int_{-\infty}^{\infty} fx(x)dx = 1$$

Samt at alle værdierne af pdf er større eller lig 0

```
teathed = int(fx(x), 'x', -Inf, Inf)
```

teathed = $3k$

```
k = solve(teathed==1, k)
```

k =

$$\frac{1}{3}$$

3) Skitsér tæthedsfunktionen og angiv navnet på fordelingsfunktionen.

```
clearvars
syms x
k = 1/3
```

k = 0.3333

```
Fx(x) = piecewise(2 >= x, 0, 2 < x <= 5, k * x -2/3, 5 < x, 1);
fx(x) = diff(Fx(x));

fplot(fx(x), [-4, 13],'-r','LineWidth',2)

title('Tæthedsfunktion')
xlabel('x')
ylabel('y')
```

Fordelingsfunktionen er uniform fordelt

4) Brug $F_X(x)$ til at beregne sandsynligheden $\Pr(x \geq 3)$. Antag at $k = \frac{1}{3}$.

```
Pr=1-Fx(3)
```

```
Pr =
```

$$\frac{2}{3}$$

5) Bestem forventningsværdien og variansen af X udfra $f_X(x)$. Angiv desuden hvilken formler, der bruges til at bestemme værdierne. Antag at $k = \frac{1}{3}$.

Forventningsværdien givet ved

$$E\;|X| = \int x \cdot f(x)\,\mathrm{d}x$$

```
Ex=int(x*fx(x),'x',-Inf,Inf)
```

```
Ex =
```

$$\frac{7}{2}$$

Variance er givet ved

$$Var(x) = E\;|x^2| - E\;|X|^2$$

```
var = int(x^2*fx(x),'x',-Inf,Inf)-(Ex^2)
```

```
var =
```

$$\frac{3}{4}$$

Opgave 2: Stokastiske Processer

En kontinuer stokastisk process er givet ved:

$$X(t) = w + 4$$

Hvor w er normalfordelt efter $w \sim N(5,1)$.

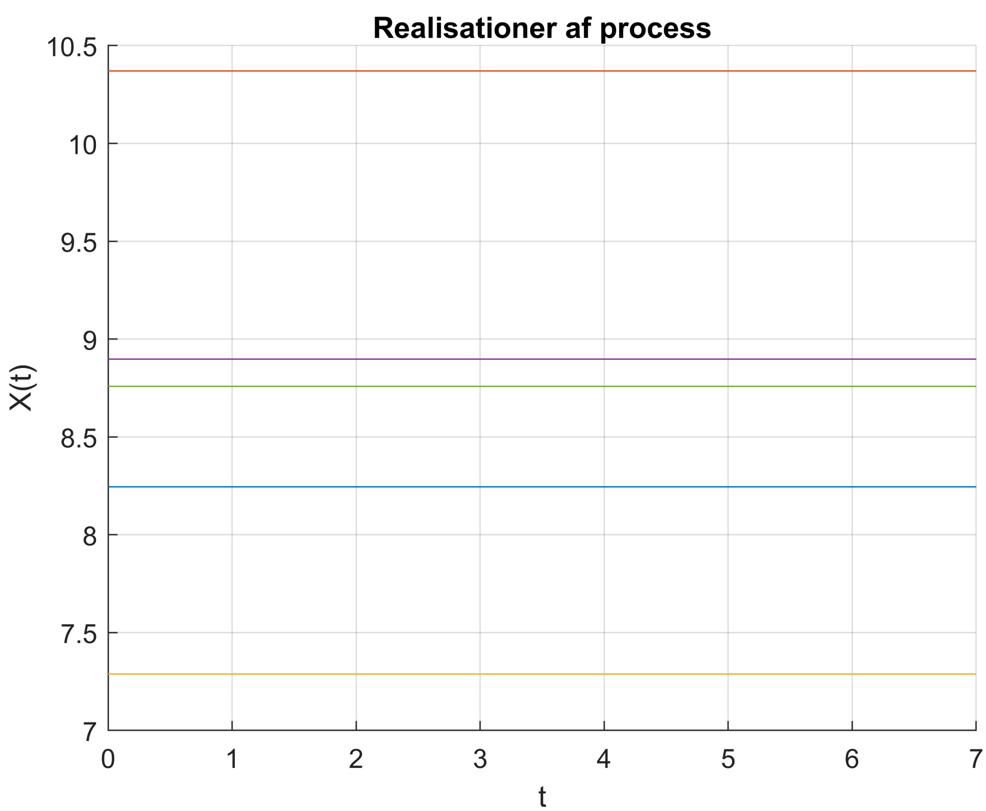
1) Skitsér fem realisationer af processen $X(t)$ mellem $t \in [0;7]$. Brug en Gauss-generator, det kan evt. være matlabs indbyggede generator, *randn()*. Angiv desuden hvordan de fem realisationer er fremkommet.

Normal fordelt - Middelværdi = 5, varians = 1

```
clearvars
close all
% Antal realisationer
N = 5;
t_start = 0;
t_end = 7;
interval = t_end-t_start;
mean = 5; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 N+1])
```

```
w = 1×6
    4.2451    6.3703    3.2885    4.8978    4.7586    5.3192
```

```
t = t_start:1:t_end;
figure(1)
hold on
grid
xlabel("t"); ylabel("X(t)"); title("Realisationer af process");
for i=1:N
    y = w(i)+4;
    fplot(y,[0 7])
end
```



2) Bestem ensemble middelværdien og ensemble variansen for processen $X(t)$.

Vi kender middelværdien for normalfordelingen W.

```
ensembleMean = mean+4
```

```
ensembleMean = 9
```

Vi kender variansen for norfordelingen W

$$Var(x(t)) = Var(w) + Var(k)$$

```
ensembleVar = variance + (4^2-4^2)
```

```
ensembleVar = 1
```

3) Udvælg én af de fem realisationer, og bestem middelværdien og variansen for denne realisation.

```
firstsRealisation = w(1)+4
```

```
firstsRealisation = 8.2451
```

```
Ex = firstsRealisation
```

```
Ex = 8.2451
```

```
% Varians = E[X^2] - E[X]^2
Ex2 = firstsRealisation^2
```

```
Ex2 = 67.9812
```

```
var = Ex2-Ex^2
```

```
var = 0
```

4) Angiv om processen $X(t)$ er WSS (stationær i den brede forstand) og om den er ergodisk. Begrund dine svar.

Da processen er uafhængig af tiden, og at middelværdien og variansen ikke ændrer sig, så kan vi konstanter at processen er stationær.

Processen er ikke ergodisk, da en realisation ikke siger noge om middelværdien, og variansen af hele prossen.

Opgave 3: Sandsynlighedsregning

Hændelse A er, at en gravid fødte en pige i 2012.

Hændelse B er, at hun fødte en dreng.

Hændelse C er, at hun fødte et barn, der vejer over 4000g.

20,2% af alle nyfødte drenge vejede i 2012 over 4000g. 12,8% af nyfødte piger vejede i 2012 over 4000g.

1) Hvis der blev født 29.785 drenge og 28.131 piger i 2012, hvad er sandsynligheden for hændelse A?

```
piger = 28131
```

```
piger = 28131
```

```
drenge = 29785
```

```
drenge = 29785
```

```
total = piger + drenge
```

```
total = 57916
```

```
Pr_a=piger/(total)
```

```
Pr_a = 0.4857
```

2) Hvad er den totale sandsynlighed for hændelse C?

Total sandsynlighed er givet ved

$$\begin{aligned}\Pr(A) &= \Pr(A \cap B) + \Pr(A \cap \bar{B}) \\ &= \Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})\end{aligned}$$

```
Pr_b= 1-Pr_a
```

```
Pr_b = 0.5143
```

```
Pr_C_A = 0.128
```

```
Pr_C_A = 0.1280
```

```
Pr_C_B = 0.202
```

```
Pr_C_B = 0.2020
```

```
Pr_C = ((Pr_C_A*Pr_a)+(Pr_C_B*Pr_b))
```

```
Pr_C = 0.1661
```

3) Hvad var sandsynligheden for at den fødende fik en pige, hvis det oplyses, at hendes barn vejede over 4000g ved fødslen?

Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

```
(Pr_C_A*Pr_a)/Pr_C
```

```
ans = 0.3744
```

Opgave 4: Statistik

Vi måler højden på studerende i en klasse, der består af 19 kvinder og 35 mænd. Højderne antages at være normalfordelt. Middelværdien for kvinder i klassen er $\widehat{\mu_1} = 1{,}68m$, med en estimeret varians på ${s_1}^2 = 0{,}10$. Middelværdien for mænd i klassen er $\widehat{\mu_2} = 1{,}78m$, med en estimeret varians på ${s_2}^2 = 0{,}20$.

1) Opstil et hypotese test, for at bestemme om middelværdien af mænd og kvinder i klassen er den samme.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

$H_1 = \mu 1 \neq \mu 2$

2) Angiv hvilken statistisk test du vil udføre for at teste hypotesen.

Begrund dit svar.

Da variansen er ukendt, og dataen er normalt fordelt, så bruges der en t-test på forskellen af middelværdierne.

3) Estimér forskellen i middelværdierne δ og standard afvigelsen σ for forskellen.

Estimator

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$$

```
middel= abs(1.68-1.78)
```

```
middel = 0.1000
```

Standard deviation er givet ved:

$$\sigma = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1) * s_2^2}{n_1+n_2-2}}$$

```
n1 = 19;
n2 = 35;
s1_2= 0.1;
s2_2 = 0.2;
standard=sqrt(((n1-1)*s1_2+(n2-1)*s2_2)/(n1+n2-2))
```

```
standard = 0.4067
```

4) Anvend en t-test til hypotese test af din hypotese. Kan NULL hypotesen afvises med et signifikansniveau på 0,05? Begrund dit svar.

$$t = \frac{\hat{\delta}}{s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

```
t = middel/(standard*sqrt((1/n1)+(1/n2)))
```

```
t = 0.8629
```

P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right)$$

```
pval = 2*(1-tcdf(abs(t),(n1+n2)-2))
```

```
pval = 0.3921
```

Da pval > 0.05 så fejler vi med at afvise null hypotesen.

5) Opstil og find 95% konfidens intervallet for δ. Angiv hvilken formel, der er brugt.

Lower bound:

$$\delta_- = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Upper bound:

$$\delta_+ = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
t0=tinv(1-0.05/2,(n1+n2-2))
```

```
t0 = 2.0066
```

```
koefMinus = middel-t0*standard*sqrt((1/n1)+(1/n2))
```

```
koefMinus = -0.1325
```

```
koefPositiv = middel+t0*standard*sqrt((1/n1)+(1/n2))
```

```
koefPositiv = 0.3325
```

```
clearvars
clc
close all
```

Opgave 1: Stokastiske Variable

En diskret stokastisk variabel X har følgende tæthedsfunktion (pmf):

x	-3	0	2	4	7	10	12
$f_X(x)$	k	k	k	k	k	k	k

1) For hvilken værdi af k er $f_X(x)$ en gyldig tæthedsfunktion? Begrund svaret.

```
clearvars
```

For at det er en gyldig tæthedsfunktion er følgende givet:

$$\sum_x fx(x) = 1$$

```
syms k
x = [-3 0 2 4 7 10 12];
k = solve(k*length(x)== 1 , k)
```

k =

$$\frac{1}{7}$$

2) Skitsér tæthedsfunktionen.

```
fx = [k k k k k k k];
plot(x,fx)

title('Tæthedsfunktion')
xlabel('x')
ylabel('y')
```

3) Bestem forventningsværdien og variansen af X udfra $f_X(x)$. Angiv hvilke formler, der bruges til at finde værdierne. Antag at $k = \frac{1}{7}$.

Ved forventningsværien gælder følgende:

$$\boldsymbol{E}|\boldsymbol{x}| = \sum_{x} x * fx(x)$$

```
Ex= sum(x.*fx)
```

Ex =

$$\frac{32}{7}$$

Variance er givet ved

$$Var(x) = E\ |x^2| - E\ |X|^2$$

```
Ex2= sum(x.^2.*fx)
```

Ex2 = 46

```
var = Ex2 - Ex^2
```

var =

$$\frac{1230}{49}$$

4) Beregn sandsynlighederne $\Pr(x \geq 2)$ og $\Pr(x > 2)$. Antag at $k = \frac{1}{7}$.

```
Pr1 = sum(fx(x>=2))
```

Pr1 =

$$\frac{5}{7}$$

```
Pr2 = sum(fx(x>2))
```

Pr2 =

$\frac{4}{7}$

5) Bestem fordelingsfunktionen (cdf) for X. Angiv desuden mellemregninger. Antag at $k = \frac{1}{7}$.

$$Fx(x) = \int_{nedre\ værdi}^{x} x\, dx + tidl.\ led$$

```
syms x
Fx(x) = piecewise( x < -3, 0, ...
              -3 <= x < 0, 1/7, ...
              0 <= x < 2, 2/7, ...
              2 <= x < 4, 3/7, ...
              4 <= x < 7, 4/7, ...
              7 <= x < 10, 5/7, ...
              10 <= x < 12, 6/7, ...
              12 <= x, 7/7)
```

Fx(x) =

$$\begin{cases} 0 & \text{if } x < -3 \\ \frac{1}{7} & \text{if } x \in [-3, 0) \\ \frac{2}{7} & \text{if } x \in [0, 2) \\ \frac{3}{7} & \text{if } x \in [2, 4) \\ \frac{4}{7} & \text{if } x \in [4, 7) \\ \frac{5}{7} & \text{if } x \in [7, 10) \\ \frac{6}{7} & \text{if } x \in [10, 12) \\ 1 & \text{if } 12 \le x \end{cases}$$

Opgave 2: Stokastiske Processer

En kontinuer stokastisk process er givet ved:

$$X(t) = w(t)$$

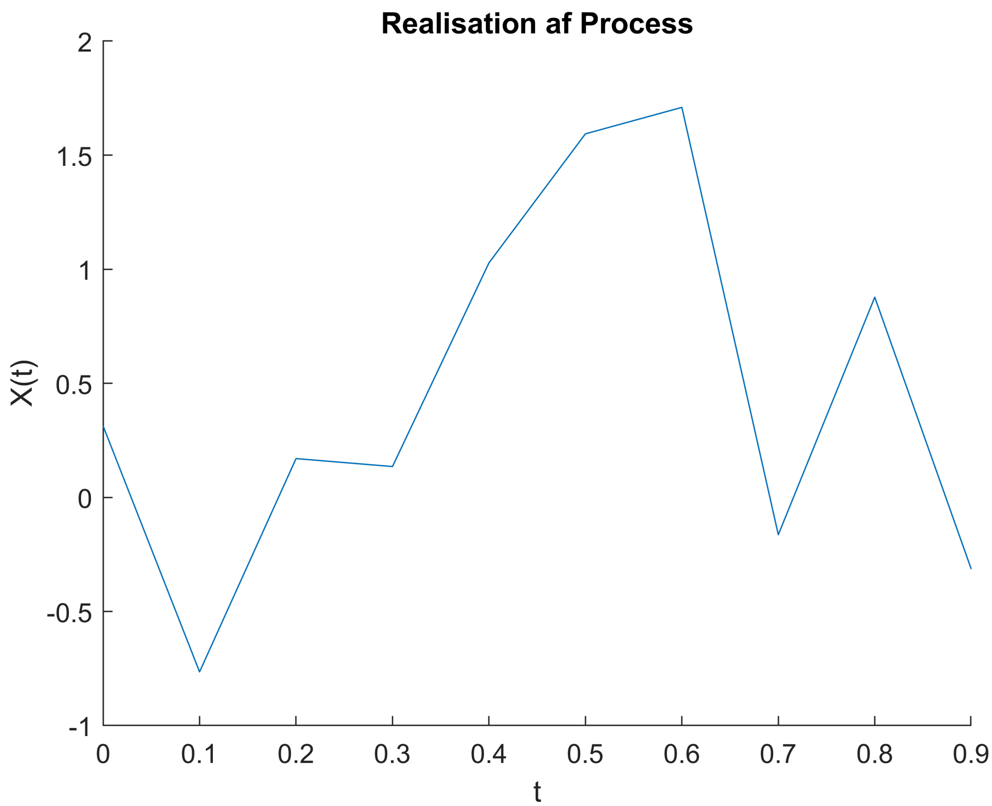
Hvor $w(t)$ er i.i.d. (uafhængig og ens fordelt) og normalfordelt efter $w(t) \sim N(t,1)$.

1) Skitsér én realisation af processen $X(t)$, hvor den er samplet til tiderne: $t = [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$. Brug en Gauss-generator, f.eks. *randn*() i matlab. Angiv desuden hvordan realisationen er fremkommet.

```
close all
t = 0:0.1:0.9;
mean = t; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 length(t)])
```

```
w = 1×10
    0.3129   -0.7649    0.1699    0.1351    1.0277    1.5933    1.7093   -0.1637 ⋯
```

```
figure(1)
hold on;
xlabel("t"); ylabel("X(t)"); title("Realisation af Process");
plot(t, w)
```



2) Bestem ensemble middelværdien og ensemble variansen for processen
$X(t)$.

Ensemble middelværdi ved w(t)~N(t,1) er t pga normalfordelingsfunktionen,

0 ved den givne stokastiske process

```
EnsembleMean = t + 0
```

```
EnsembleMean = 1×10
         0    0.1000    0.2000    0.3000    0.4000    0.5000    0.6000    0.7000 ⋯
```

ensemble variansen ved w(t)~N(t,1) er 1 pga normalfordelingsfunktionen

og +0 ved den givne stokastiske process

```
EnsembleVar = 1 + 0
```

```
EnsembleVar = 1
```

3) Hvad forventer du den tidslige middelværdi af en vilkårlig realisation af
$X(t)$ i et tidsinterval $t = [0; 100]$ vil blive, begrund dit svar?

```
t = 0:1:100;
mean = t; variance = 1; sigma = sqrt(variance);
w = normrnd(mean, sigma, [1 length(t)])
```

```
w = 1×101
   -1.1135    0.9932    3.5326    2.2303    4.3714    4.7744    7.1174    5.9109 ···
```

```
clear mean
mean(w)
```

```
ans = 49.9215
```

4) Angiv om processen $X(t)$ er WSS (stationær i den brede forstand) og om den er ergodisk. Begrund dine svar.

Da processen er afhængig af tiden, og at middelværdien og variansen ændrer sig over tid, så kan vi konstanter at processen ikke er stationær.

Da processen ikke er stationær, så kan den heller ikke være ergodisk.

5) Opstil ligningen til bestemmelse af Autocorrelationen $R_{X(t_1)X(t_2)}(t_1 = 1, t_2 = 2)$ og udregn værdien.

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$$

```
t1 = 1; t2=2;
E_T1 = 1 + 0;
E_T2= 2 + 0;
E_T1 * E_T2
```

```
ans = 2
```

Opgave 3: Sandsynlighedsregning

Et studie viser at hvis et barn på 14 er flyttet mere end én gang på et år, vil barnet med en sansynlighed på 0,06 begå alvorlig kriminalitet indenfor de næste 10 år. For børn, der flyttede én eller færre gange på et år, var sandsynligheden 0,03.

31% af børnene i studiet tilhørte gruppen, der var flyttet mere end en gang.

1) Hvad er den totale sandsynlighed for at et af børnene i studiet begik alvorlig kriminalitet indenfor de næste 10 år?

Hændelse A : Barn er flyttet mere end en gang

Hændelse B : Begår kriminalitet

```
format
Pr_B_A = 0.06;
Pr_B_NOT_A = 0.03;
Pr_a = 0.31;
Pr_not_a = 1-Pr_a;
```

Total sandsynlighed er givet ved

$$\begin{aligned}\Pr(A) &= \Pr(A \cap B) + \Pr(A \cap \bar{B}) \\ &= \Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})\end{aligned}$$

```
Pr_b = Pr_B_A*Pr_a+Pr_B_NOT_A*Pr_not_a
```

```
Pr_b = 0.0393
```

2) Hvis et barn fra studiet har begået alvorlig kriminalitet indenfor de 10 år studiet rakte sig over, hvad er sandsynligheden for at barnet tilhørte gruppen, der havde flyttet mere end én gang?

Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

```
SandsynlighedTilhorteGruppe = (Pr_B_A*Pr_a)/Pr_b
```

```
SandsynlighedTilhorteGruppe = 0.4733
```

Opgave 4: Statistik

I et studie af tandhvalers forventede levetid, registrerede man dødsalderen på individuelle tandhvaler. Der blev i studiet registreret 10 hvaler, der var døde i fangenskab, og 10 hvaler, der var døde i det fri.

Død i det fri (alder i år)	Død i Fangenskab (alder i år)
50	7
43	2
11	1
35	3
7	15
62	6
70	14
67	1
25	5
1	9

```
fri = [50 43 11 35 7 62 70 67 25 1 ];
fange = [7 2 1 3 15 6 14 1 5 9];
```

1) Opstil en NULL og Alternativ hypotese, for at bestemme om middelværdien af de to grupper er den samme.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

$H_1 = \mu 1 \neq \mu 2$

2) Bør testen, der udføres, være parret eller uparret? Begrund dit svar.

Testen være udføres uparrat, da dataens subject (fangenskab eller fri) er to forskellige ting.

3) Estimér middelværdierne for begge grupper.

```
friMean = mean(fri)
```

```
friMean = 37.1000
```

```
fangeMean = mean(fange)
```

```
fangeMean = 6.3000
```

4) Estimér varianserne for begge grupper, samt den samlede varians (pooled variance).

Sample variance for 1: $s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2$

Sample variance for 2: $s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2$

```
S2_fange = (1/(length(fange)-1)) * sum((fange-fangeMean).^2)
```

```
S2_fange = 25.5667
```

```
S2_fri = (1/(length(fri)-1)) * sum((fri-friMean).^2)
```

```
S2_fri = 648.7667
```

Pooled variance

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right)$$

```
PooledVar = ...
    (((length(fange)-1) * S2_fange) +...
    ((length(fri)-1) * S2_fri)) /...
    (length(fange)+length(fri)-2)
```

```
PooledVar = 337.1667
```

5) Anvend en uparret t-test til hypotese test af din hypotese. Kan NULL hypotesen afvises med et signifikansniveau på 0,05? Begrund dit svar.

$$t = \frac{\hat{\delta}}{s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

```
t = abs(friMean-fangeMean)/(sqrt(PooledVar)*sqrt((1/length(fri))+(1/length(fange))))
```

```
t = 3.7507
```

P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right)$$

```
pval = 2*(1-tcdf(abs(t),(length(fri)+length(fange))-2))
```

```
pval = 0.0015
```

Da pval er mindre end 0.05 så afvises hypotesen.

6) Opstil og find 95% konfidens intervallet for forskellen i middelværdierne. Angiv hvilken formel, der er brugt.

Lower bound:

$$\delta_{-} = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Upper bound:

$$\delta_{+} = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$tinv(0.975, n - 2)$$

```
n1 = length(fri)
```

```
n1 = 10
```

```
n2 = length(fange)
```

```
n2 = 10
```

```
middel = abs(friMean-fangeMean)
```

```
middel = 30.8000
```

```
t0=tinv(0.975,(n1+n2-2))
```

```
t0 = 2.1009
```

```
koefMinus = middel-t0*sqrt(PooledVar)*sqrt((1/n1)+(1/n2))
```

```
koefMinus = 13.5477
```

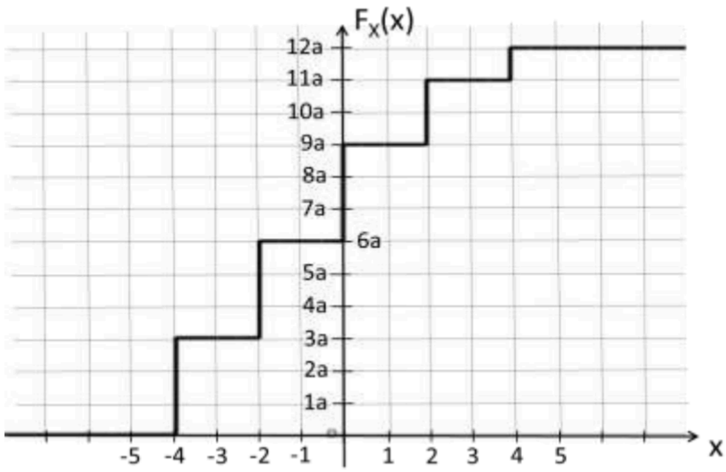
```
koefPositiv = middel+t0*sqrt(PooledVar)*sqrt((1/n1)+(1/n2))
```

```
koefPositiv = 48.0523
```

```
clearvars
clc
close all
```

Opgave 1

En diskret stokastisk variabel X har følgende fordelingsfunktion (cdf) $F_X(x)$:



a) Bestem a, så $F_X(x)$ er en gyldig fordelingsfunktion.

```
clearvars
```

$$Fx(x) = 12a = 1 => a = \frac{1}{12}$$

b) Bestem og tegn tæthedsfunktionen (pmf) $f_X(x)$.

```
fx = [3/12 3/12 3/12 2/12 1/12];
x = [-4 -2 0 2 4];

figure(1);
bar(x,fx);

sum(fx)
```

```
ans = 1
```

c) Bestem middelværdien for X.

Ved forventningsværdien gælder følgende:

$$\boldsymbol{E}|\boldsymbol{x}| = \sum_{x} x * fx(x)$$

```
middel = sum(x.*fx)
```

```
middel = -0.8333
```

d) Bestem variansen for X.

$$\boldsymbol{var(x) = E\left|x^2\right| - E|x|^2}$$

```
var = sum(x.^2.*fx)-middel^2
```

```
var = 6.3056
```

Opgave 2

I juni måned (30 dage) regner det i gennemsnit 20% af dagene i den første halvdel af måneden og 30% af dagene i den sidste halvdel af måneden.

a) Hvor mange dage regner det i gennemsnit i juni måned?

hændelse a. Regner i starten af juni.

Hændelse b. Regner i slutnigen af juni.

```
Pr_R_Forst = 0.2;
Pr_R_Sidst = 0.3;
Pr_Forst = 0.5;
Pr_Sidst = 0.5;
Juni_Dage = 30;
total_prop = Pr_R_Forst*Pr_Forst + Pr_R_Sidst * Pr_Sidst
```

```
total_prop = 0.2500
```

```
gennemsnitRegn = 30*total_prop
```

```
gennemsnitRegn = 7.5000
```

b) Hvis vi oplever en dag med regn i juni måned, hvad er så sandsynligheden for at vi er i den sidste halvdel af måneden?

Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

```
Pr_R_Sidst*Pr_Sidst/total_prop
```

```
ans = 0.6000
```

c) Hvad er sandsynligheden for at det regner **højst** 1 dag i den første halvdel af juni?

$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

```
Pr_IkkeRegn_forst = 1-Pr_R_Forst
```

```
Pr_IkkeRegn_forst = 0.8000
```

```
(15/1)*Pr_R_Forst^1*Pr_IkkeRegn_forst^14
```

```
ans = 0.1319
```

Opgave 3

En kontinuert stokastisk proces X(t) er givet ved:

$$X(t) = (-1)^n + W$$

hvor W er i.i.d. Gaussisk fordelte stokastiske variableW ~ $\mathcal{N}(0; 0{,}25)$, og n uafhængigt kan antage værdierne 0 og 1 med lige stor sandsynlighed.

a) Skitser 3 realisationer af processen X(t) i intervallet 0 ≤ t ≤ 5. Angiv hvordan de 3 realisationer er opnået.

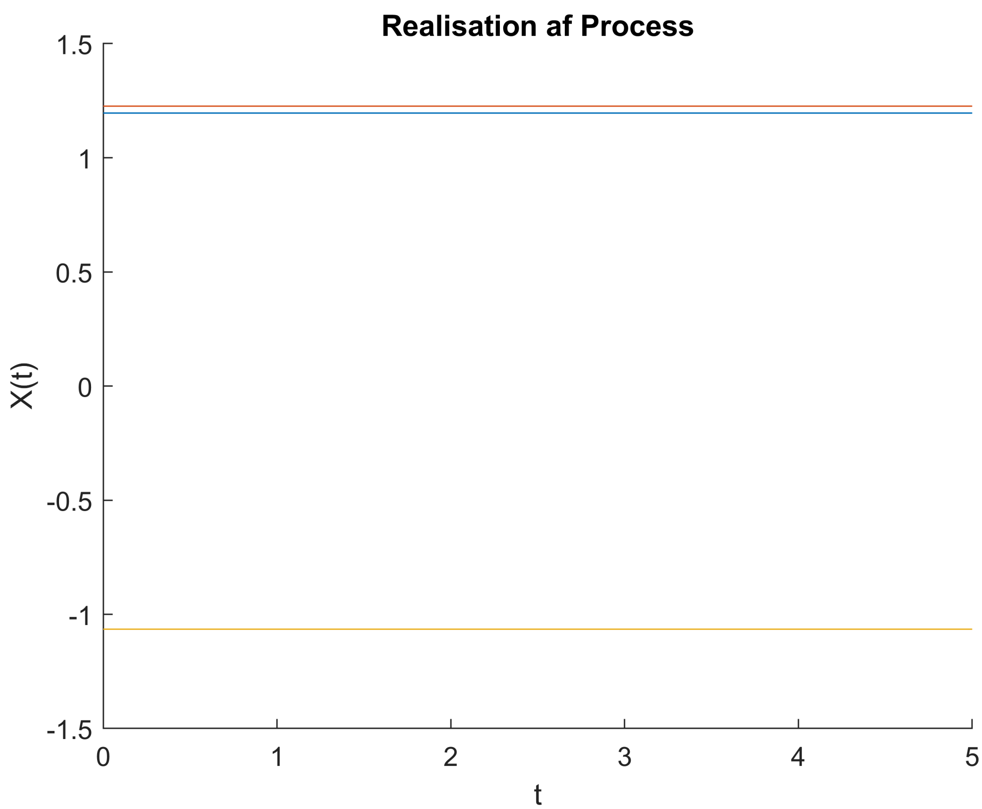
```
clearvars
close all
x = 0:1:5;
realisationer = 3;
% Først få værdierne til n
n = randi([0, 1], 1, 3)
```

```
n = 1×3
     0     0     1
```

```
% Ud fra W~N(0; 0,25) kan vi udlede:
mean = 0; variance = 0.25; sigma = sqrt(variance);
clear mean
% Få værdierne til w
w = randn(1, 3) * sigma;
figure(1)
hold on
xlabel("t"); ylabel("X(t)"); title("Realisation af Process");
```

```
for i = 1:realisationer
    y = (-1)^n(i) + w(i);
    plot(x, ones(1,length(x))*y);

end
```



b) Bestem middelværdien og variansen for én af realisationerne.

```
mean = mean(y)
```

```
mean = -1.0651
```

```
variane = var(y)
```

```
variane = 0
```

c) Bestem ensemble middelværdien og variansen for processen X(t).

$E|X(t)| = E(-1^{n}) + E(w) = -1^{0} * \Pr(0) + (-1) * \Pr(1) + E(w)$

```
middel = (-1)^0*(1/2)+(-1)^1*(1/2)+0
```

```
middel = 0
```

```
middel2 = ((-1)^0)^2*(1/2)+((-1)^1)^2*(1/2)+0
```

```
middel2 = 1
```

Varians er givet ved

$$var(x) = E|x^2| - E|x|^2 + Var(w)$$

```
variance = middel2-middel^2+0.25
```

```
variance = 1.2500
```

d) Angiv om processen er WSS (stationær i den brede forstand), og om den er ergodisk. Svarene skal begrundes.

Da processen er uafhængig af tiden, og at middelværdien og variansen ikke ændrer sig over tid, så kan vi konstanter at processen er stationær.

Processen er ikke ergodisk da pmf ikke kan bestemmes udfra en realisation

Opgave 4

En kvalitetskontrol måler præcisionen af to forskellige typer gps'er. For begge typer blev målt afvigelsen mellem deres faktiske position ($d_{faktisk}$) og gps'ens angivelse (d_{gps}):

$$d_i = |d_{i,gps} - d_{i,faktisk}|$$

Det kan antages at afvigelserne er normalfordelte.
Der er testet 10 gps'er af type 1 og 12 gps'er af type 2.
For type 1 var middelafvigelsen $\widehat{\mu_1} = 5{,}21\,m$ med en estimeret varians $s_1^2 = 1{,}33\,m^2$.
For type 2 var middelafvigelsen $\widehat{\mu_2} = 4{,}18\,m$ med en estimeret varians $s_2^2 = 0{,}89\,m^2$.

a) Opstil en hypotese test for at bestemme om middelafvigelserne for de to typer gps'er er den samme.

Null hypothesis. Brugt som testen af ingen forskel.

$H_0 = \mu 1 = \mu 2$

Alternanive hypotese

$H_1 = \mu 1 \neq \mu 2$

b) Estimer forskellen i middelværdierne $\hat{\delta}$ for de to typer.

Estimator

$$\hat{\delta} = \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2\right)$$

```
middel= abs(5.21-4.18)
```

```
middel = 1.0300
```

c) Estimer variansen $\hat{s}^2$ for forskellen mellem de to typer.

$$s^2 = \frac{1}{n_1+n_2-2}\left((n_1-1)s_1^2 + (n_2-1)s_2^2\right)$$

```
n1 = 10;
n2 = 12;
s1_2= 1.33;
s2_2 = 0.89;
standard2=1/(n1+n2-2)*((n1-1)*s1_2+(n2-1)*s2_2)
```

```
standard2 = 1.0880
```

d) Anvend en t-test til test af din hypotese. Kan NULL-hypotesen afvises med et signifikantniveau på 0,05? Svaret skal begrundes.

$$t = \frac{\hat{\delta}}{s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

```
t = middel/(sqrt(standard2)*sqrt((1/n1)+(1/n2)))
```

```
t = 2.3062
```

P-value

$$2 \cdot \left(1 - t_{cdf}(|t|, n_1 + n_2 - 2)\right)$$

```
pval = 2*(1-tcdf(abs(t),(n1+n2)-2))
```

```
pval = 0.0319
```

Da p er mindre end 0.05 så kan vi afvise hypotese 0.

e) Bestem 95% konfidens intervallet for forskellen i middelværdier δ.

Lower bound:

$$\delta_{-} = (\bar{x}_1 - \bar{x}_2) - t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Upper bound:

$$\delta_{+} = (\bar{x}_1 - \bar{x}_2) + t_0 \cdot s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
t0=tinv(1-0.05/2,(n1+n2-2))
```

```
t0 = 2.0860
```

```
koefMinus = middel-t0*sqrt(standard2)*sqrt((1/n1)+(1/n2))
```

```
koefMinus = 0.0984
```

```
koefPositiv = middel+t0*sqrt(standard2)*sqrt((1/n1)+(1/n2))
```

```
koefPositiv = 1.9616
```