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 =

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

```
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 $w(t) \sim U(-2, -1)$.

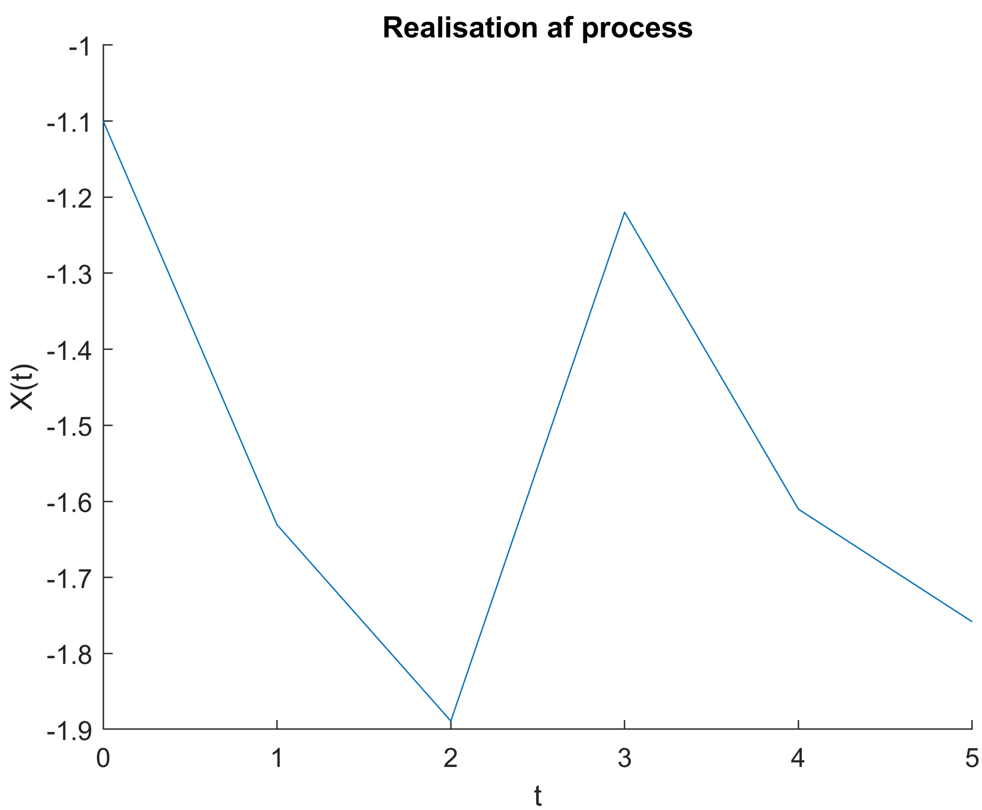
1) Den stokastiske process $X(t)$ er samplet hvert sekundt, skitsér 6 samples fra 0 - 5 s af én realisation. Angiv hvorledes realisationen er fremkommet, brug en tilfældighedsgenerator, f.eks. rand() i matlab.

```
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

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

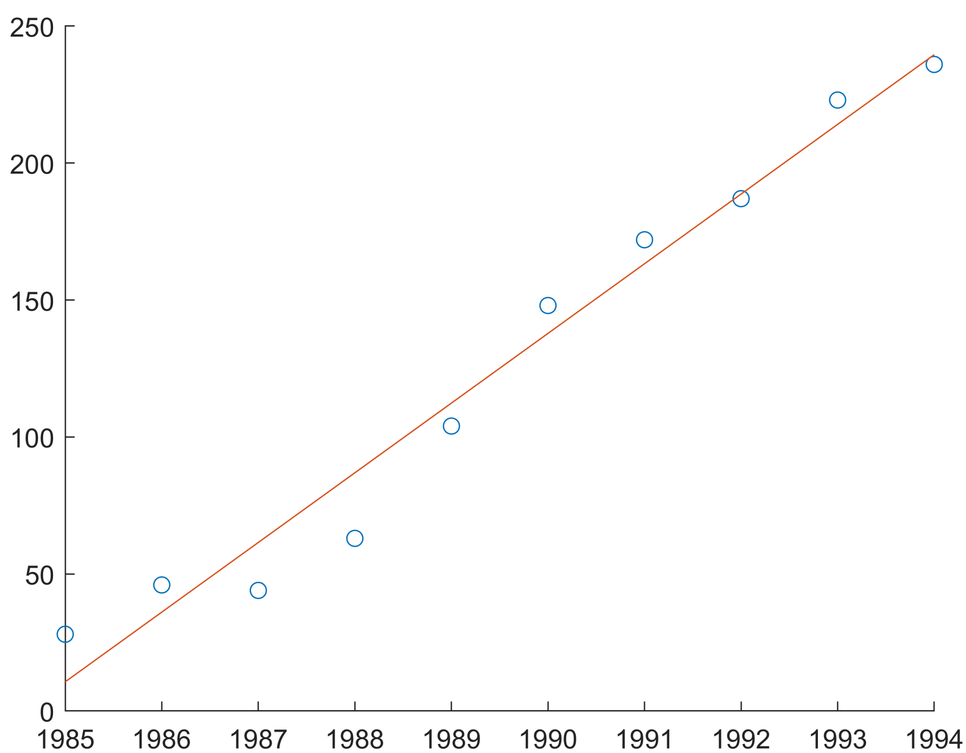
```
%Parameter estimat hældning
Beta = sum((antal - emperiskeMiddel) .* (years - yearsMean)) / sum((years - yearsMean).^2)
```

```
Beta = 25.4364
```

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

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

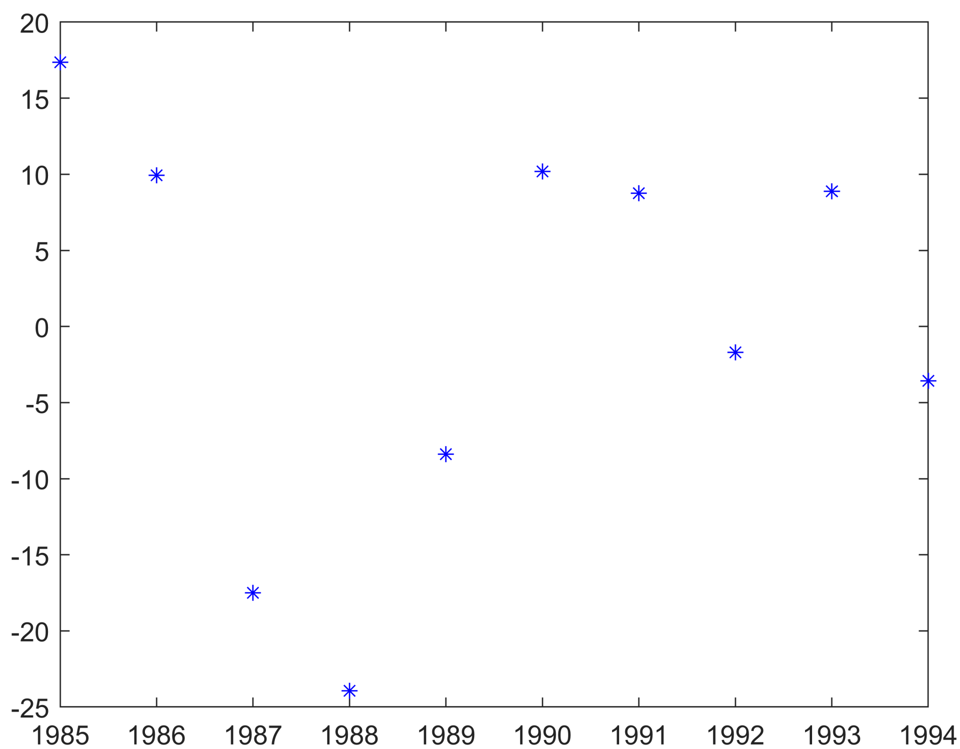
```
linearModel = Alpha + years * Beta;
figure(2)
hold on
scatter(years, antal)
plot(years,linearModel)
```



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.