Opgave 1: Stokastiske Variable

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

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

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

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

Løst i mathcad prime 4.0

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

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

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

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

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

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

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

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

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

Opgave 2: Stokastiske Processer

En diskret stokastisk process er givet ved:

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

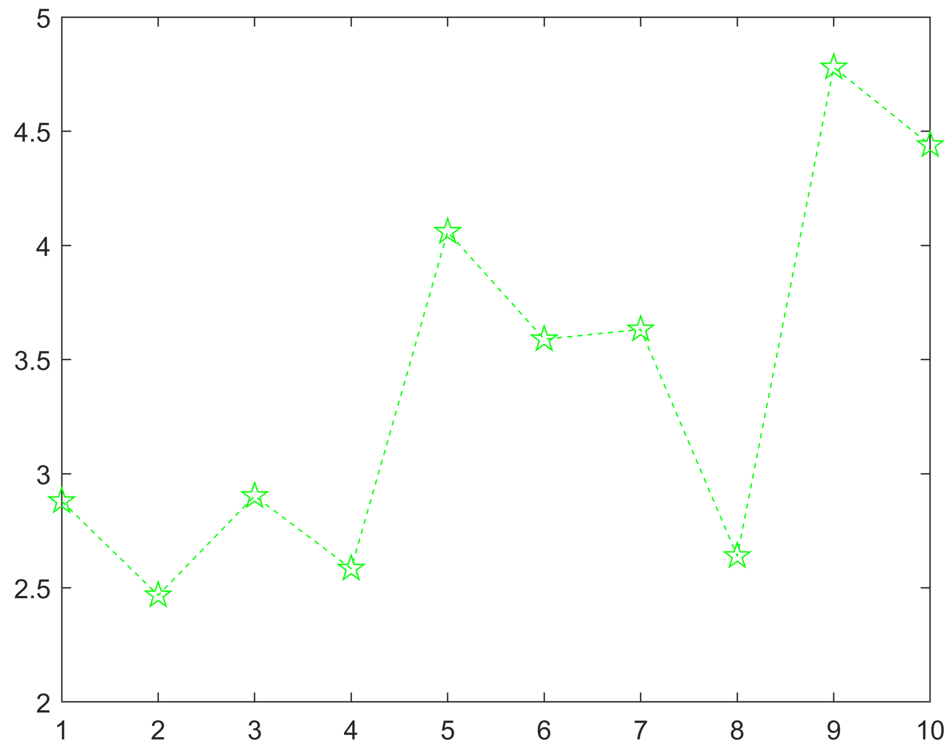
Hvor hver sample n af w er i.i.d Gaussisk fordelte stokastiske variable $w(n) \sim N(0,1)$.

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

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

```
n =
     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 =
   0.134615384615385
```

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 = 
   0.134615384615385
```

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

```
PrBgivetA = 
   0.117647058823529
```

```
PrA_B = PrBgivetA * Pr_b
```

```
PrA_B = 
   0.015837104072398
```

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 =
  30.200000000000003
```

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

```
yearsMean =
        1991
```

```

```

Benytter følgende formel for at finde skæringen og hældningen

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

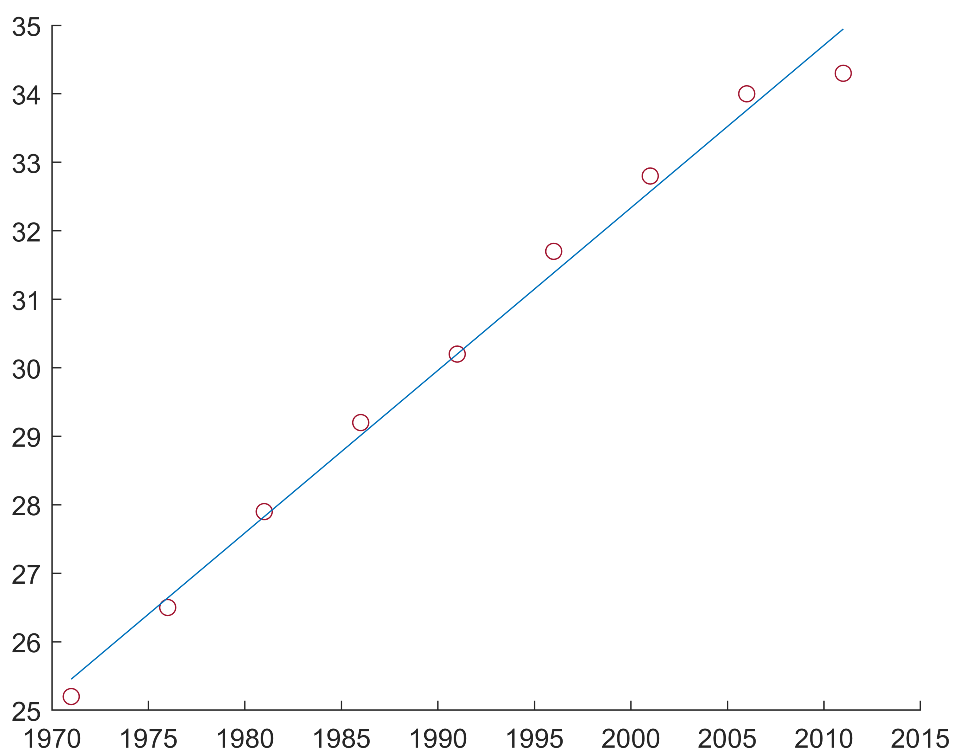
```
%Parameter estimat hældning
Beta = sum((Alder - alderMean) .* (Years - yearsMean)) / sum((Years - yearsMean).^2)
```

```
Beta =
   0.237333333333333
```

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

```
Alpha =
    -4.423306666666666e+02
```

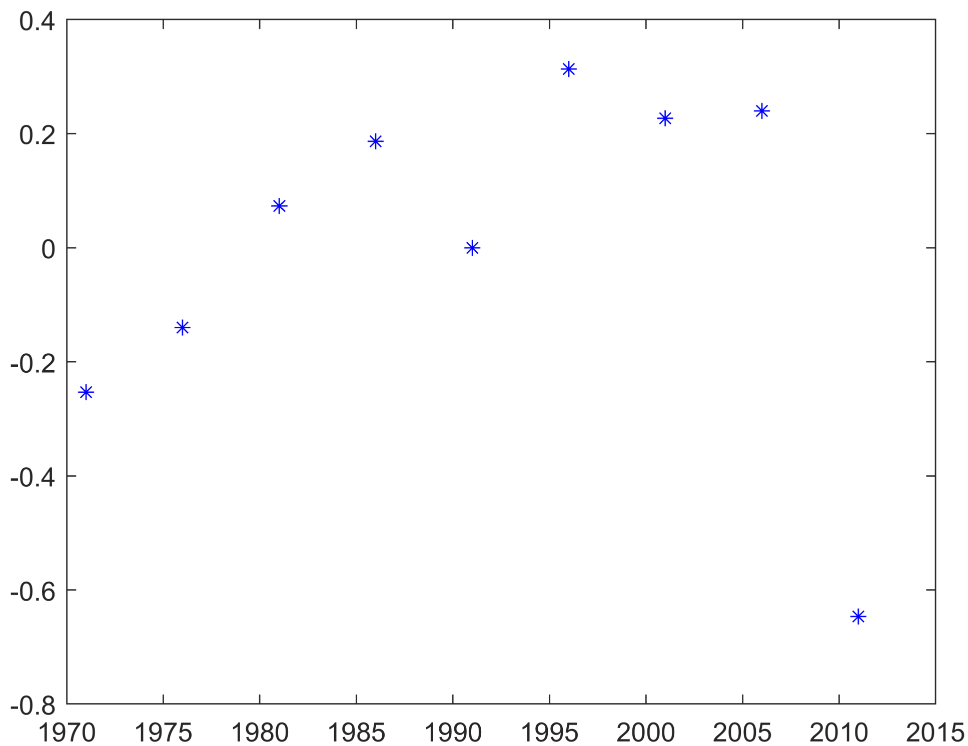
```
linearModel = Alpha + Years * Beta;
figure(2)
hold on
scatter(Years, Alder)
plot(Years,linearModel)
```



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.