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)
```

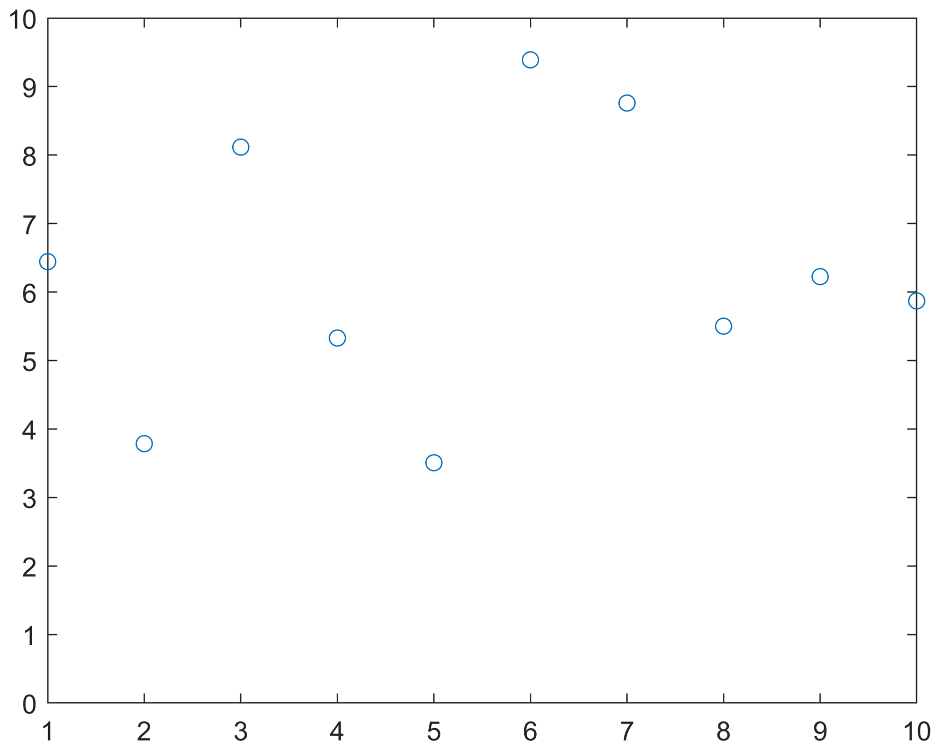
```
fxy =
       1/2            0              1/2
```

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
```

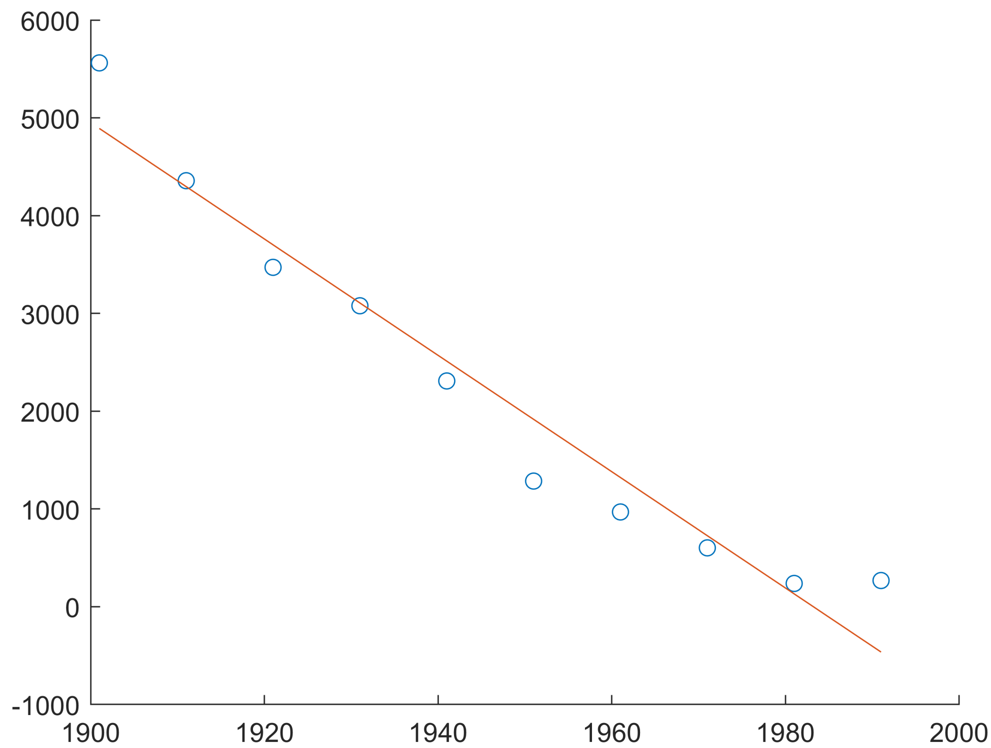
```


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

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

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

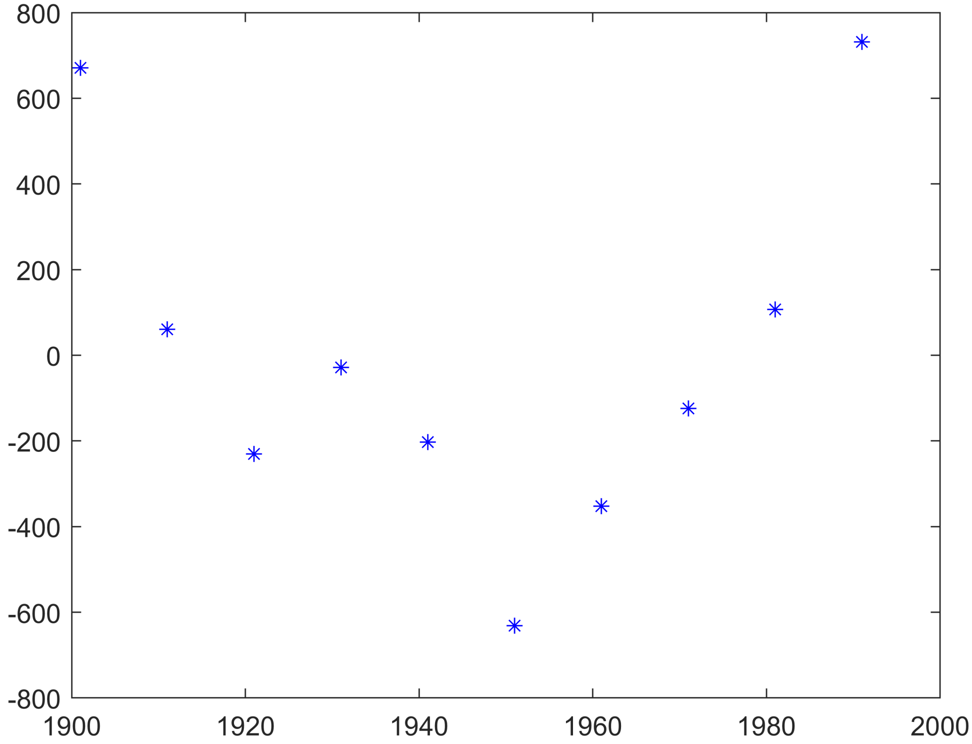
```
plot(years,linearModel)
```



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
```

```
t0 = tinv(0.974,(n-2))
```

```
t0 = 2.2809
```

95% konfidensintervallet kan derefter beregnes med følgende to formler:

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

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

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

```
ConfMinus = -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.