

## Contents

Stokastiske variabler.....	2
CDF til PDF funktionen:.....	2
PMF til CDF funktionen:.....	3
PDF til CDF funktionen:.....	3
For hvilken værdi af $k$ er $f_X(x)$ en gyldig tæthedsfunktion?(PMF) Begrund svaret. ....	3
For hvilken værdi af $k$ er $f_X(x)$ en gyldig tæthedsfunktion ud fra PMF.....	4
For hvilken værdi af $k$ er $f_X(x)$ en gyldig tæthedsfunktion ud fra PDF.....	4
For hvilken værdi af $k$ er $f_X(x)$ en gyldig tæthedsfunktion ud fra PDF.....	4
For hvilken værdi af $k$ er $f_X(x)$ en gyldig tæthedsfunktion ud fra PDF.....	5
Vis den marginale tæthedsfunktion (PDF) .....	5
Brug $F_X(x)$ til at beregne sandsynligheden $\Pr(x \geq 3)$ . $k=1/3$ .....	5
Brug $F_X(x)$ til at beregne sandsynligheden $\Pr(x < 0,4)$ og $\Pr(0,1 \leq x < 0,4)$ . $k = 1/e$ .....	6
Brug $F_X(x)$ til at beregne sandsynligheden $\Pr(x < 0)$ .....	6
Brug $f_X(x)$ (PDF) til at finde forventningsværdien $E[X]$ og variansen $\sigma_X^2$ for $X$ .....	6
Find Forventningsværdien $E[X]$ og variansen $\sigma_X^2$ for $X$ . Fra $f_X(x)$ (PDF) Angiv formler der benyttes.....	7
Brug $f_X(x)$ (PMF) til at finde forventningsværdien $E[X]$ og standard afvigelsen $\sigma_X$ for $X$ . ....	7
Find forventningsværdien $E[g(X=x)]$ , Funktionen defineret som $g(X=x)=3 \cdot x^2$ .....	8
Bestem forventningsværdien og variansen af $X$ ud fra $f_X(x)$ (PMF). Angiv hvilke formler, der bruges til at finde værdierne. $K=1/7$ .....	8
Stokastiske Processer .....	8
Bestem ensemble middelværdien og ensemble variansen for processen $X(n)$ (Datasæt) .....	8
Bestem ensemble middelværdien og ensemble variansen for processen.....	9
Sandsynlighedsregning .....	9
Komplementærreglen .....	9
Bayes Regel.....	9
Betinget Sandsynlighed .....	9
Total Sandsynlighed.....	9
Foreningshændelse .....	9
Uafhængighed .....	10
Kombinationer.....	10
Eksempel 1 alm. Sandsynlighedsregning inkl. den totale sandsynlighed.....	10
Poisson fordeling .....	11
Beregn sandsynligheden .....	11
Passer understående antagelse, samt lav en hypotesetest eller lign. Brug et signifikansniveau på 5% .....	11

Statistik .....	12
Opstil en NULL og Alternativ hypotese, for at bestemme om middelværdien af de to grupper er den samme. ....	12
Estimér middelværdierne for begge grupper.....	12
Estimér varianserne for begge grupper, samt den samlede varians (pooled variance). ....	13
Estimér forskellen i middelværdierne $\delta$ og standard afvigelsen $\sigma$ for forskellen. ....	13
Anvend en t-test til hypotese test af din hypotese. Kan NULL hypotesen afvises med et signifikansniveau på 0,05? Begrund dit svar.....	13
Opstil og find 95% konfidens intervallet for $\delta$ . Angiv hvilken formel, der er brugt. ....	14
Eksempel på Matlab udregning.....	14
Plot data fra tabellen, bestem skæringen og hældningen af den lineære model.....	14
Lav en residualtegning.....	15
Beregn et 95 % konfidensinterval for hældningen.....	16
Begrund, at det med rimelighed kan antages, at forskellen i målingerne er normalfordelt (Parret data)...	17

## Stokastiske variabler

Find tæthedssfunktionen (PDF) for følgende fordelingsfunktion (CDF):

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

CDF til PDF funktionen:

$$f_X(x) = \frac{d F_X(x)}{dx}$$

Differentier samtlige værdier:

$$2 \geq x = \frac{d 0}{dx} = 0, \quad 2 < x \leq 5 = \frac{d k \cdot x - \frac{2}{3}}{dx} = k, \quad 5 < x = \frac{d 1}{dx} = 0$$

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

Find fordelings funktionen ud fra følgende tæthedsfunktion (PMF) og antag  $k=1/8$

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

PMF til CDF funktionen:

$$\sum_{x=-1}^x f_X(x), \quad \text{note: husk tidl. led}$$

Find  $F_X(x)$  for de forskellige værdier af  $x$ :

$$F_X(-1) = \sum_{x=-1}^{-1} k = \frac{1}{8}, \quad F_X(1) = \sum_{x=-1}^1 k + \frac{3}{4} = \frac{7}{8}, \quad F_X(7) = \sum_{x=-1}^7 k + \frac{3}{4} + \frac{1}{8} = 1$$

Derved er Fordelingsfunktionen givet ved:

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{8} & -1 \leq x < 1 \\ \frac{7}{8} & 1 \leq x < 7 \\ 1 & 7 \leq x \end{cases}$$

Find fordelingsfunktion (CDF) følgende tæthedssfunktionen (PDF):

$$f(x) = \begin{cases} \frac{1}{3} & 0 < x < 1 \\ \frac{2}{7} * x^2 & 1 < x < 2 \end{cases}$$

PDF til CDF funktionen:

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

Find  $F_X(x)$  for de forskellige værdier af  $x$ :

$$F_X\left(\frac{1}{3}\right) = \int_0^{\frac{1}{3}} \frac{1}{3} \, dx = \frac{x}{3}, \quad F_X\left(\frac{2}{7} * x^2\right) = \int_1^x \frac{2}{7} * x^2 \, dx + \frac{1}{3} = \frac{2x^3 + 5}{21}$$

Derved er Fordelingsfunktionen givet ved:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{3} & 0 \leq x < 1 \\ \frac{2x^3 + 5}{21} & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

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

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

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

For at det er en gyldig tæthedsfunktion er følgende givet:

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

Da  $F_X(x)$  er ens for hele tæthedsfunktionen, må den være uniformt fordelt. Da der er tale om en tæthedsfunktion skal summen af alle  $k$  være lig 1, derfor kan  $k$  kun være lig:

$$1 = k * 7 \Rightarrow k = 1/7$$

For hvilken værdi af  $k$  er  $f_X(x)$  en gyldig tæthedsfunktion ud fra PMF

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

For at det er en gyldig tæthedsfunktion er følgende givet:

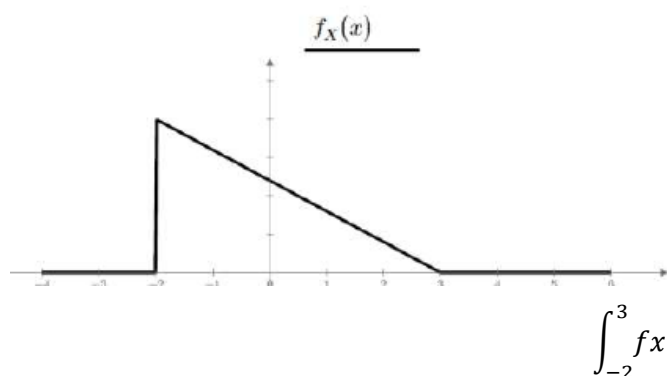
$$\sum_x f_X(x) = 1$$

Vi isolerer  $k$

$$k = 1 - \sum_x f_X(x), \quad 1 - \frac{1}{8} - \frac{3}{4} = \frac{1}{8}, \quad k = \frac{1}{8}$$

For hvilken værdi af  $k$  er  $f_X(x)$  en gyldig tæthedsfunktion ud fra PDF

På skitsen er  $f_X(3) = 0$ . find tæthedsfunktionen for  $f_X(-2)=k$



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

For at

$f_X(x)$  er en gyldig tæthedsfunktion gælder følgende:

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

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

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

For hvilken værdi af  $k$  er  $f_X(x)$  en gyldig tæthedsfunktion ud fra PDF

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

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

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

Samt at alle værdierne af pdf er større eller lig 0

$$\int_{-\infty}^1 k * e^x dx + \int_1^{\infty} 0 dx = e * k \rightarrow e * k = 1 \rightarrow \frac{1}{e}$$

For hvilken værdi af  $k$  er  $f_X(x)$  en gyldig tæthedsfunktion ud fra PDF

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

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

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

Derved kan  $k$  findes:

$$\int_2^5 k dx = 1 \rightarrow 3 * k = 1 \rightarrow k = \frac{1}{3}$$

Vis den marginale tæthedsfunktion (PDF)

Lad  $X$  og  $Y$  være kontinuerte stokastiske variable med følgende simultane tæthedsfunktion (PDF):

$$f(x, y) = \begin{cases} \frac{3}{16}(x^3 + y^2 + 1) & \text{hvis } -1 \leq x \leq 1 \text{ og } -1 \leq y \leq 1 \\ 0 & \text{ellers} \end{cases}$$

Den marginale tæthedsfunktionen er givet ved:

$$f_X(X, Y) = \int_{-\infty}^{\infty} f_X(X, Y) d(x, y)$$

Derved kan jeg vise den marginale tæthedsfunktion for  $X$  og  $Y$

$$f_X(x) = \int_{-1}^1 \frac{3}{16}(x^3 + y^2 + 1) dy = \frac{3}{8}x^3 + \frac{1}{2}$$

$$f_Y(y) = \int_{-1}^1 \frac{3}{16}(x^3 + y^2 + 1) dx = \frac{3}{8}(y^2 + 1)$$

Brug  $F_X(x)$  til at beregne sandsynligheden  $\Pr(x \geq 3)$ .  $k=1/3$

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

Følgende gælder

$$Pr(x \geq 3) = Fx(3), \quad Pr(x \geq 3) = 1 - Pr(x < 3)$$

$$Pr(x \geq 3) = 1 - \left(k * x - \frac{2}{3}\right) = \frac{2}{3}$$

Brug  $Fx(x)$  til at beregne sandsynligheden  $Pr(x < 0,4)$  og  $Pr(0,1 \leq x < 0,4)$ .  $k = 1/e$

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

Følgende gælder for første led:

$$Pr(x < 0,4) = Fx(0,4), \quad Pr(x < 0,4) = k * e^x = 0,55$$

Følgende gælder for andet led:

$$Pr(0,1 \leq x < 0,4) = Pr(x \leq 0,4) - Pr(x < 0,1)$$

$$Pr(x < 0,1) = Fx(0,1), \quad Pr(x < 0,1) = k * e^x = 0,41$$

$$Pr(0,1 \leq x < 0,4) = 0,55 - 0,41 = \mathbf{0,14}$$

Brug  $Fx(x)$  til at beregne sandsynligheden  $Pr(x < 0)$

$$A = -\frac{2}{25}$$

$$B = \frac{6}{25}$$

$$C = \frac{16}{25}$$

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

Følgende gælder:

$$Pr(x < 0) = Fx(0), \quad Pr(x < 0,4) = \frac{A}{2} * x^2 + B * x + C = \frac{16}{25}$$

Brug  $fx(x)$ (PDF) til at finde forventningsværdien  $E[X]$  og variansen  $\sigma_x^2$  for  $X$

$$A = -\frac{2}{25}$$

$$B = \frac{6}{25}$$

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

Ved forventningsværdien gælder følgende:

$$E|x| = \int_{-\infty}^{\infty} x * f(x) dx \rightarrow \int_{-2}^3 x * \left(-\frac{2}{25} * x + \frac{6}{25}\right) dx = -\frac{1}{3}$$

Ved varians af  $x$  gælder følgende:

$$var(x) = E|x^2| - E|x|^2 =$$

$$\int_{-\infty}^{\infty} x^2 * f_X(x) dx - \left(\frac{7}{2}\right)^2 \int_2^3 x^2 \left(-\frac{2}{25} * x + \frac{6}{25}\right) dx - \left(\frac{1}{3}\right)^2 \rightarrow$$

$$1,5 - \left(\frac{1}{3}\right)^2 = 1.39$$

Find Forventningsværdien  $E[X]$  og variansen  $\sigma_x^2$  for  $X$ . Fra  $f_X(x)$ (PDF) Angiv formler der benyttes

$$k = \frac{1}{3}$$

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

Ved forventningsværdien gælder følgende:

$$E|x| = \int_{-\infty}^{\infty} x * f_X(x) dx \rightarrow \int_2^5 \frac{1}{3} * x dx = \frac{7}{2}$$

Tabelopslag for uniform fordeling

$$E|x| = \frac{a+b}{2} = \frac{5+2}{2} = \frac{7}{2}$$

Ved varians af  $x$  gælder følgende:

$$\text{var}(x) = E|x^2| - E|x|^2 = \int_{-\infty}^{\infty} x^2 * f_X(x) dx - \left(\frac{7}{2}\right)^2 \rightarrow \int_2^5 \frac{1}{3} * x^2 dx - \left(\frac{7}{2}\right)^2 \rightarrow 13 - \left(\frac{7}{2}\right)^2 = \frac{3}{4}$$

Tabelopslag for uniform fordeling

$$E|x| = \frac{(b-a)^2}{12} = \frac{(5-2)^2}{12} = \frac{3}{4}$$

Brug  $f_X(x)$  (PMF) til at finde forventningsværdien  $E[X]$  og standard afvigelsen  $\sigma_x$  for  $X$ .

$$k = \frac{1}{8}$$

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

Ved forventningsværdien gælder følgende:

$$E|x| = \sum_x x * f_X(x) \rightarrow -1 * k + 1 * \frac{3}{4} + 7 * \frac{1}{8} = 1,5$$

Ved standard afvigelsen gælder følgende:

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

$$E|x^2| = \sum_x x^2 * f_x(x) \rightarrow -1^2 * k + 1^2 * \frac{3}{4} + 7^2 * \frac{1}{8} = 7$$

$$\sigma_x = \sqrt{7 - 1,5^2} = 2,18$$

Find forventningsværdien  $E[g(X=x)]$ , Funktionen defineret som  $g(X=x)=3*x^2$

$$k = \frac{1}{8}$$

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

Hvis

$$g(x) = 3 * x^2, \quad \text{finder man } E[g(x)], \quad \text{hvor } k = \frac{1}{8}$$

$$E[3 * x^2] \rightarrow 3 * E|x^2| \rightarrow 3 * 7 = 21$$

Bestem forventningsværdien og variansen af X ud fra  $f_x(x)$ (PMF). Angiv hvilke formler, der bruges til at finde værdierne.  $k=1/7$

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

Ved forventningsværdien gælder følgende:

$$E|x| = \sum_x x * f_x(x) \rightarrow -3 * k + 0 * k + 4 * k + 7 * k + 10 * k + 12 * k = 4,57$$

Varians er givet ved

$$var(x) = E|x^2| - E|x|^2$$

$$E|x^2| = \sum_x x^2 * f_x(x) \rightarrow -3^2 * k + 0^2 * k + 4^2 * k + 7^2 * k + 10^2 * k + 12^2 * k = 46$$

$$var(x) = 46 - 4,57^2 = 25$$

## Stokastiske Processer

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

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

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

Ved middelværdien gælder følgende:

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

Ved variansen gælder følgende:

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



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

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

Bestem ensemble middelværdien og ensemble variansen for processen

**En kontinuer stokastisk process er givet ved:**

$$X(t) = w + 5$$

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

Da middel af  $w(n) = 5$  er ensemble mean lig summen af de to middelværdier fra hhv.  $w(n)$  og de 5 der bliver lagt til. Ensemble middelværdien er derfor følgende:

$$E|X(t)| = E|w| + E|5| = 10$$

De 5 der bliver lagt til er ligegyldige for variansen da:

$$var[Y] = var[a * x + b] = a^2 * var(x), \quad \text{Da vi intet } a \text{ og } b \text{ har, går de ud, og får følgende:}$$

$$var|X(t)| = var|w| + var|5| = 1$$

## Sandsynlighedregning

### Komplementærreglen

$$\Pr(\overline{A}) = 1 - \Pr(A)$$

$$\Pr(\overline{B}|A) = 1 - \Pr(B|A)$$

$$\Pr(\overline{B}|\overline{A}) = 1 - \Pr(B|\overline{A})$$

### Bayes Regel

$$\Pr(A \cap B) = \Pr(A | B) * P(B)$$

Så længe man er konsekvent i forhold til udskiftninger af konstanter er det okay:

$$\Pr(A \cap B) = \Pr(B | A) * P(A)$$

$$\Pr(\overline{A} \cap B) = \Pr(B | \overline{A}) * P(\overline{A})$$

$$\Pr(A \cap \overline{B}) = \Pr(\overline{B} | A) * P(A)$$

$$\Pr(\overline{A} \cap \overline{B}) = \Pr(\overline{B} | \overline{A}) * P(\overline{A})$$

### Betinget Sandsynlighed

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

### Total Sandsynlighed

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \overline{B})$$

### Foreningshændelse

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

## Uafhængighed

To hændelser er uafhængige hvis:

$$\Pr(A \cap B) = P(A) * \Pr(B)$$

## Kombinationer

### Sorteret med tilbagelægning

Et objekt tages fra en samling af  $n$  objekter. Objektet lægges tilbage hver gang. Eksperimentet gentages  $k$  gange. Rækkefølges på objektet betyder noget. Antallet af kombinationer er:

$$n^k$$

### Sorteret uden tilbagelægning

Et objekt tages fra en samling af  $n$  objekter. Objektet lægges ikke tilbage hver gang. Eksperimentet gentages  $k$  gange. Rækkefølges på objektet betyder noget. Antallet af kombinationer er:

$$\frac{n!}{(n - k)!}$$

### Usorteret med tilbagelægning

Et objekt tages fra en samling af  $n$  objekter. Objektet lægges tilbage hver gang. Eksperimentet gentages  $k$  gange. Rækkefølges på objektet betyder ikke noget. Antallet af kombinationer er:

$$\frac{(n + k - 1)!}{k! * (n - k)!}$$

### Usorteret uden tilbagelægning

Et objekt tages fra en samling af  $n$  objekter. Objektet lægges ikke tilbage hver gang. Eksperimentet gentages  $k$  gange. Rækkefølges på objektet betyder ikke noget. Antallet af kombinationer er:

$$\frac{n!}{k! * (n - k)!}$$

## Eksempel 1 alm. Sandsynlighedsregning inkl. den totale sandsynlighed

Hændelse A: Det er en pige

Hændelse B: Det er en dreng

Hændelse C: Et barn på over 4000g.

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

Der blev født **29.785 drenge og 28.131 piger**.

Sandsynligheden for A:

$$\Pr(A) = \frac{28.131}{28.131 + 29.785} = 0,4857 * 100 = 48,57\%$$

Total sandsynlighed for C, først findes sandsynligheden for B:

$$\Pr(B) = 1 - \Pr(A) = 0,5143$$

Total sandsynlighed for C:

$$\Pr(C) = \Pr(C|A) \Pr(A) + \Pr(C|B) \Pr(B)$$

$$\Pr(C) = 0.128 * 0.4957 + 0.202 * 0.5143 = 0.1661$$

Sandsynligheden for det er en pige, hvis det oplyses, at hendes barn vejede over 4000g.

$$\Pr(A) = 0,4857$$

$$\Pr(C) = 0.1661$$

$$\Pr(A|C) = \frac{\Pr(A) \Pr(C|A)}{\Pr(C)} = \frac{0,4857 * 0.128}{0.1661} = 0,3744$$

## Poisson fordeling

I et bestemt lyskryds antager man, at der i gennemsnit passerer 180 biler per time i morgentrafikken. På grund af vejarbejde forventer man, at der vil opstå trafikprop længere fremme, hvis mere end 5 biler passerer lyskrydset over en periode på et minut.

Lad  $X$  være en stokastisk variabel, som angiver antallet af biler, som passerer lyskrydset per minut. Det kan med rimelighed antages, at  $X$  følger en Poisson-fordeling

$$X \sim \text{poisson}(\lambda \cdot t)$$

hvor tidsparameteren antages at være

$$\lambda = \frac{x}{t} = \frac{\text{antal biler}}{\text{tid i minutter}} = \frac{180}{60} = 3 \text{ biler per minut}$$

## Beregn sandsynligheden

Sandsynligheden for at der kommer mindst 5 (over 4 biler) biler er givet ved

$$\Pr(X > 4) = 1 - F_{\text{poisson}}(4, \lambda)$$

`t = 60; %Indtast tid. Tidsinterval`

`x = 180; %Antallet af events observeret over en periode af tid t`

$$\lambda = \frac{x}{t}$$

`%Find Pr(X > 4)`

`%Pr(X > 4) = 1-Pr(<= 4)`

`%Fx(4) hvor X~poisson(lambda)`

`x = 4;`

$$\text{poiss} = 1 - \text{poisscdf}(x, \lambda) = \mathbf{0.1847}$$

Passer understående antagelse, samt lav en hypotesetest eller lign. Brug et signifikansniveau på 5%

2. Over en periode på fem hverdage, hvor man hver dag har talt, hvor mange biler, der passerer det givne lyskryds over en periode på to timer i morgentrafikken, har man observeret, at der passerer 1700 biler. Den samlede observationstid er 10 timer. Stemmer denne observation overens med ovenstående antagelse, at der passerer 180 biler i timen? Du skal lave et hypotesetest eller noget tilsvarende for at besvare denne opgave korrekt. Brug et signifikansniveau på 5%.

`λ0 = 180; % Events pr. time i gennemsnit`

`t = 10; % Indtast tid. Tidsinterval`

`x = 1700; % Antallet af events observeret over en periode af tid t`

Parameter estimat / Maximum-likelihood estimator:

$$\hat{\lambda} = \frac{x}{t}$$

Hypotese Test

$$H_0: \lambda = \lambda_0$$

$$H_1: \lambda \neq \lambda_0$$

Test-size

$$z = \frac{x - \lambda * t}{\sqrt{\lambda * t}} = -2.357$$

Approximate p-værdi

$$p_{val\_app} = 2 * \text{abs}(1 - \text{normcdf}(\text{abs}(z))) = 0.0184$$

Da p-værdien er under 0.05 kan hypotesen forkastes!

Observationerne passer ikke til antagelsen at der passerer 180 biler i timen.

## Statistik

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

Nulhypotese:

$$H_0: \hat{\mu}_1 = \hat{\mu}_2$$

Alternativ hypotese:

$$H_1: \hat{\mu}_1 \neq \hat{\mu}_2$$

Estimer middelværdierne for begge grupper.

For at finde den empiriske middelværdi gælder følgende:

$$\mu_x = \frac{\sum_{i=1}^{n_x} (x_i)}{n_x}$$

Dette blev udregnet i matlab:

```
x1=[50 43 11 35 7 62 70 67 25 1];
x2 = [7 2 1 3 15 6 14 1 5 9];
n = 10;
vpa(sum(x1/n))
vpa(sum(x2/n))
```

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

$$\mu_{x1} = \frac{\text{sum}[x1]}{n} = 37.1$$

$$\mu_{x2} = \frac{\text{sum}[x2]}{n} = 6.3$$

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

For at finde variansen gælder følgende:

$$\sigma_x^2 = \frac{\sum_{i=1}^{n_x} (x_i - \mu_x)^2}{n_x - 1}$$

Dette blev regnet i Matlab:

$$S2_{Fri} = \left( \frac{\text{sum}((x1 - \text{avg1}).^2)}{\text{length}(x1) - 1} \right) \rightarrow \sigma_{x1}^2 = 648.77$$

$$S2_{Fange} = \left( \frac{\text{sum}((x2 - \text{avg2}).^2)}{\text{length}(x2) - 1} \right) \rightarrow \sigma_{x2}^2 = 25.57$$

Derved kan pooled variance udregnes:

$$\frac{((\text{length}(x2) - 1) * S2_{Fange}) + ((\text{length}(x1) - 1) * S2_{Fri})}{\text{length}(x2) + \text{length}(x1) - 2} = 337.16$$

Estimér forskellen i middelværdierne  $\delta$  og standard afvigelsen  $\sigma$  for forskellen.

Hvor  $\hat{\mu}$  er givet ved:

$$= \frac{(\sum_{i=1}^{n_x} (x_i))}{n_x - 1}$$

Standard deviation er givet ved:

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

Dette udregnes i matlab:

```
Kvinde = 19;
Kmu = 1.68;
Ksigma_i_anden = 0.10;
Mand = 35;
Mmu = 1.78;
Msigma_i_anden = 0.20;

S = sqrt((1/(Kvinde+Mand-2)) * ...
          ((Kvinde-1)*Ksigma_i_anden) + ((Mand-1)*Msigma_i_anden))

sigma = S = 0.4067
```

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

T-test er givet ved:

$$\delta = \hat{\mu}_x - \hat{\mu}_y$$

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

Dette udregnes i Matlab:

$$t = \frac{\hat{\delta}}{S * \sqrt{\left(\frac{1}{Kvinde}\right) + \left(\frac{1}{Mand}\right)}}$$

$$p = 2 * (1 - tcdf(abs(t), Kvinde + Mand - 2)) = 0.3921$$

da  $p \gg 0.05$  kan i ikke afvise vores 0 hypotese.

Opstil og find 95% konfidens intervallet for  $\delta$ . Angiv hvilken formel, der er brugt.

95 % konfidensinterval er givet ved

$$\hat{\delta} = \hat{\mu}_x - \hat{\mu}_y$$

$$\delta_- = \hat{\delta} - t_0 * s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

$$\delta_+ = \hat{\delta} + t_0 * s * \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

Derved udregnes  $t_0$

$$t_0 = \text{tinv}(0.975, Kvinde + Mand - 2) = 2.0066$$

Inden konfidensintervallet kan udregnes, finder vi deltamean:

$$DeltaMean = abs(Kmu - Mmu)$$

Derved har vi alle de nødvendige værdier:

$$DeltaMean + t_0 * S * \sqrt{\left(\frac{1}{Kvinde}\right) + \left(\frac{1}{Mand}\right)} = 0.3325$$

$$DeltaMean - t_0 * S * \sqrt{\left(\frac{1}{Kvinde}\right) + \left(\frac{1}{Mand}\right)} = -0.1325$$

Eksempel på Matlab udregning

**Antal:** er dødelighed af drenge under 1 år i Danmark fra 1901 til 1991. Det er angivet i tabellen<sup>1</sup>:

<b>Antal:</b>	5562	4357	3471	3078	2309	1285	969	602	238	268
<b>År:</b>	1901	1911	1921	1931	1941	1951	1961	1971	1981	1991

Plot data fra tabellen, bestem skæringen og hældningen af den lineære model

Hældningen  $\hat{\beta}$ , i en lineær regression er givet ved følgende formel:

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

Derved har jeg udregnet hældningen i Matlab:

```
y = [5562 4357 3471 3078 2309 1285 969 602 238 268];
x = [1901 1911 1921 1931 1941 1951 1961 1971 1981 1991];
```

$$\bar{x} = \text{mean}(x);$$

$$\bar{y} = \text{mean}(y);$$

$$\hat{\beta} = \frac{\text{sum}((x - \bar{x}) * (y - \bar{y}))}{\text{sum}((x - \bar{x})^2)} = -59.500$$

Skæringen  $\hat{\alpha}$ , i en lineær regression er givet ud fra følgende formel: An

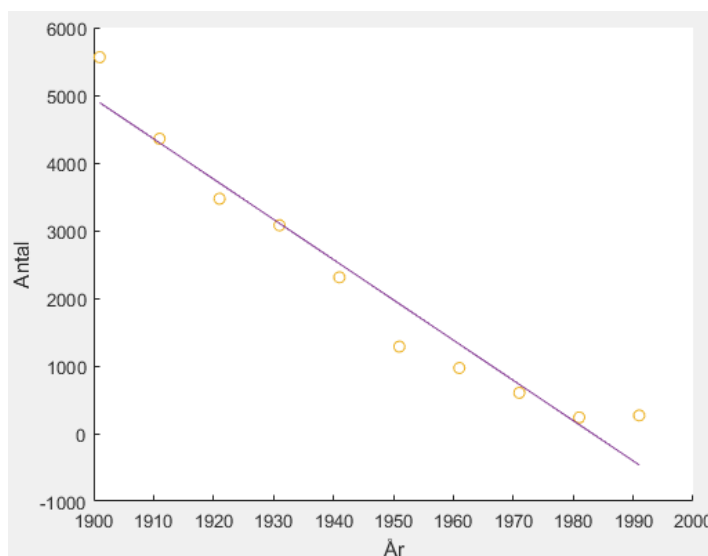
$$\hat{\alpha} = \bar{y} - \hat{\beta} * \bar{x}$$

Derved har jeg udregnet skæringen i Matlab:

$$\hat{\alpha} = \bar{y} - \hat{\beta} * \bar{x} = 1.1800 * 10^5$$

Yderligere kan jeg plotte data sættet i Matlab:

```
figure(1)
scatter(x,y)
ylabel('Antal')
xlabel('År')
hold on
plot(x,  $\hat{\alpha} + \hat{\beta} * x$ )
```



## Lav en residualtegning

Jeg kan beregne hver residualvædi ud fra følgende formel:

$$\epsilon_i = x_i - (\hat{\alpha} + \hat{\beta} * t_i)$$

Dette benyttes på den sammenholdte data med lineære regression graf 2):

$$\epsilon = y - (\hat{\alpha} + \hat{\beta} * x) =$$

e =

```
670.6000    60.6000   -230.4000   -28.4000  -202.4000  -631.4000  -352.4000  -124.4000   106.6000   731.6000
```

Dernæst kan et residual plot laves i Matlab:

```
figure(2)
hold off
r=y-(â+β*x);
scatter(x, r);
hold on
ylabel('Antal')
xlabel('År')
title('Residual')
hold on
```



Beregn et 95 % konfidensinterval for hældningen

Hvor  $s^2$  er givet ved:

$$s_r^2 = \sum_{i=1}^n \frac{(x_i - (\hat{\alpha} + \hat{\beta} * t_i))^2}{n - 2}$$

Først findes standard deviation i Matlab:

$$n = \text{length}(x) = 10 \quad \% \text{sample size}$$

$$s_r^2 = \frac{1}{n - 2} * \text{sum}((y - (\hat{\alpha} + \hat{\beta} * x)).^2) = 2.0415 * 10^5 \quad \% \text{variance}$$

$$s_r = \sqrt{s_r^2} = 451.8244 \quad \% \text{Standard afvigelse}$$

Yderligere ved jeg at  $t_0$  er givet ved:

$$t_0 = \text{tinv}(0.975, n - 2) = 2.3060$$

95% konfidensintervallet kan derefter beregnes med følgende to formler:

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

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

Derved kan jeg i Matlab udregne konfidensintervallet:

$$\hat{\beta}_- = \hat{\beta} - t_0 * s_r * \sqrt{\left( \frac{1}{\text{sum}((x - \bar{x}).^2)} \right)} = -70.9710$$

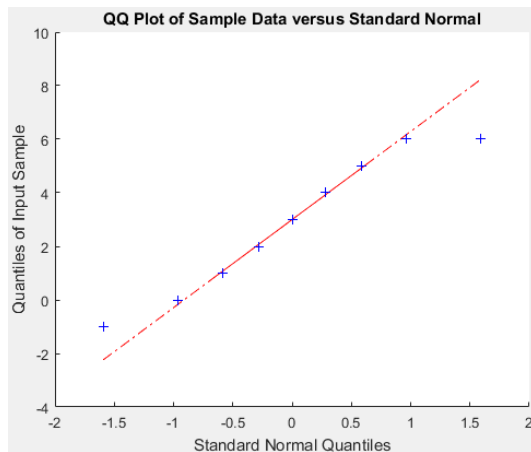
$$\hat{\beta}_+ = \hat{\beta} + t_0 * s_r * \sqrt{\left( \frac{1}{\text{sum}((x - \bar{x}).^2)} \right)} = -48.0290$$



Begrund, at det med rimelighed kan antages, at forskellen i målingerne er normalfordelt (Parret data)

Jeg benyttede matlab:

```
x4=[2 1 -1 0 6 6 5 3 4];  
qqplot(x4)
```



Det ses at datapunkterne følger den røde linje og er dermed normalfordelt