

# **OBLIGATORISK INNLEVERING 1, STA-1001 2024**

13-02-2024

# Innholdsfortegnelse

Oppgave 1: . . . . .	3
Oppgave 2: . . . . .	4
Oppgave 3: . . . . .	5
a) Om han skal velge 8 av de 10 låtene: . . . . .	5
b) Om han skal velge 8 av de 10 låtene: . . . . .	6
c) Om han skal velge 6 av de 10 låtene, men må ha med B om han har med A (om motsatt), og kan . . . . .	6
d) Om han skal velge 8 av de 10 låtene og velger tilfeldig: . . . . .	7
Oppgave 4 . . . . .	7
a) Er hendelsene $A = \text{"laks har parasitt A"}$ og $B = \text{"laks har parasitt B"}$ uavhengige? . . . . .	7
b) Regn ut sannsynligheten for at en laks har begge parasittene $P(A \cap B)$ . . . . .	7
c) Gitt at en laks har minst en parasitt, hva er sannsynligheten for at den har parasitt A? . . . . .	8
d) Hva er sannsynligheten for at minst en laks har parasitt A? . . . . .	8
Oppgave 5 . . . . .	9
a) Skisser tettheten i R. . . . .	9
b) Når er det siste tidspunktet du fremdeles kan være 90% sikker på at et batteri virker? . . . . .	11
c) Finn sannsynligheten for at et tilfeldig valgt batteri skal virke etter ett år. . . . .	11
Oppgave 6 . . . . .	12
a) Finn de marginale fordelingene (sannsynlighetstetthetene) for X og Y? . . . . .	12
b) Finn ved integrasjon av den simultane sannsynlighetstettheten . . . . .	14
Referanser . . . . .	15

## Oppgave 1:

Gjør oppgave 1.18 i boka ved bruk av R. (Om du vil slippe å skrive inn dataene ligger det ei fil karakterer.txt på Canvas.)

Bruk også R til å lage punktdiagram (funksjonen stripchart) og box-plott for dataene.

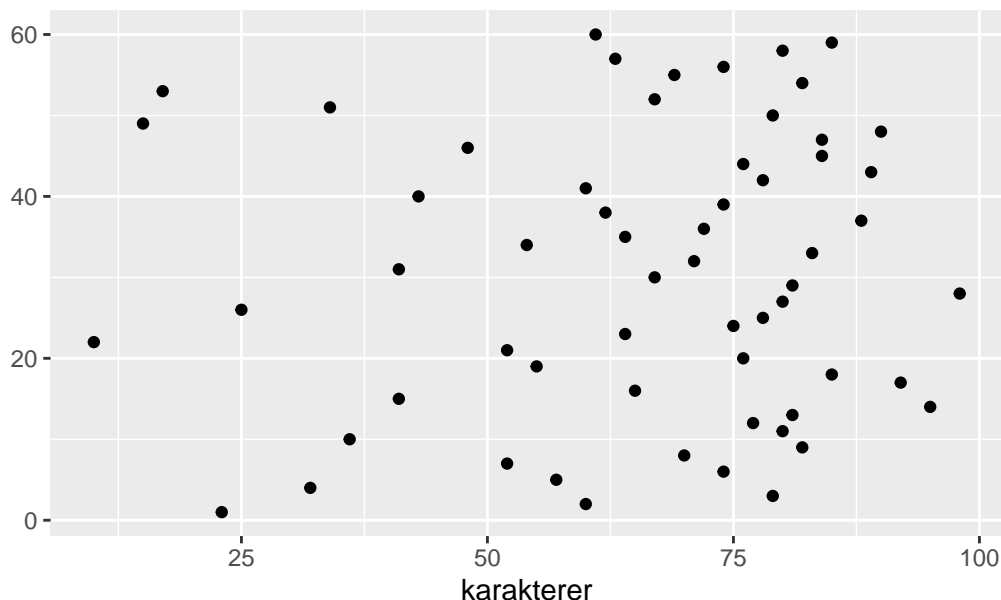
```
library(tidyverse)
library(stringr)
karakterer <- "23 60 79 32 57 74 52 70 82 36 80 77 81 95 41 65 92 85 55 76 52 10 64 75 7
karakterer <- str_replace_all(karakterer, " ", "\n")

karakterer <- read.csv(text=karakterer, header=FALSE)

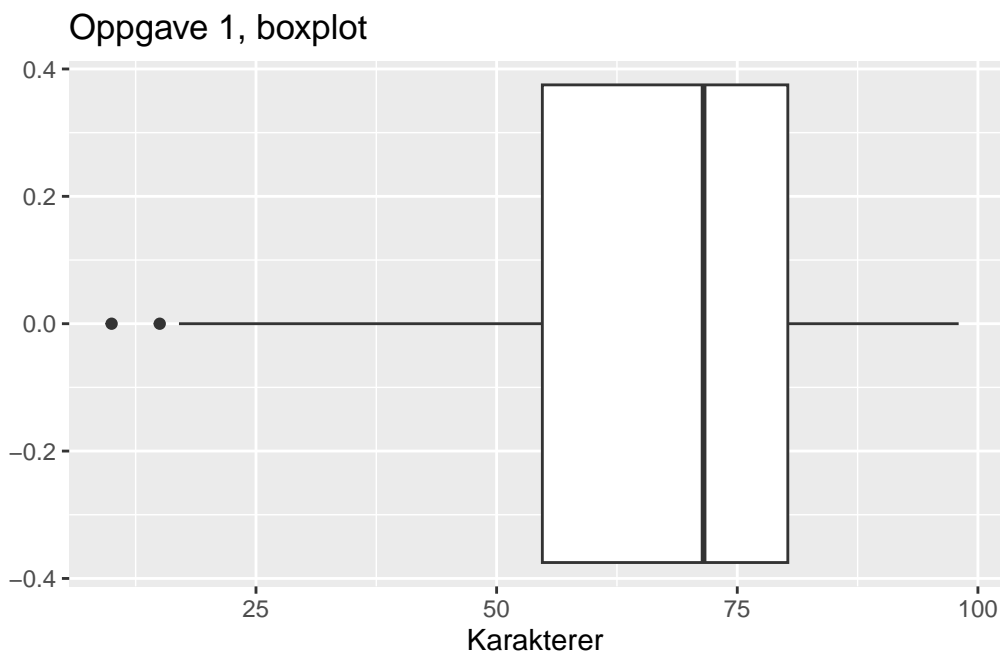
karakterer <- karakterer %>%
  rename(karakterer = V1) %>%
  mutate(index = row_number())

karakterer %>%
  ggplot(aes(x=karakterer, y=index)) +
  geom_point()+
  labs(y="", title="Oppgave 1, punkt diagram")
```

Oppgave 1, punkt diagram



```
karakterer %>%
  ggplot(aes(x=karakterer)) +
  geom_boxplot() +
  labs(title="Oppgave 1, boxplot", x="Karakterer", y="")
```



## Oppgave 2:

I boken “Sannsynlighetsregning og statistikk for høyere utdanning” på side 29 så har jeg benyttet meg av bevis 1.10 for hjelp av utregning.<sup>1</sup>

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

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

For å vise dette, starter jeg med å gange ut venstre siden av ligning 1.

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

Denne deler vi så opp:

$$\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2$$

Gitt at et aritmetisk gjennomsnitt er regnet med

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

---

<sup>1</sup>Kristensen & Wikan (2019)

så kan jeg forenkle det andre leddet ved å bruke ligning 2.

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

som da endelig gir meg

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

Setter den så inn igjen i ligning 1

$$\sum_{i=1}^n x_i^2 - 2 \cdot n \cdot \bar{x}^2 + \sum_{i=1}^n \bar{x}^2$$

Så starter jeg på det tredje leddet og gitt at  $\bar{x}$  er konstant så skriver jeg den om til

$$\sum_{i=1}^n \bar{x}^2 = n \cdot \bar{x}^2$$

Til slutt får jeg som endelig forenkles ut i ligning 3.

$$\begin{aligned} \sum_{i=1}^n x_i^2 - 2 \cdot n \cdot \bar{x}^2 + n \cdot \bar{x}^2 \\ = \sum_{i=1}^n x_i^2 - n\bar{x}^2 \end{aligned} \quad (3)$$

### Oppgave 3:

En artist skal gi ut et album. Problemet er at artisten har ti låter, og må velge bort to av dem. Han vet heller ikke hvilken rekkefølge han vil bruke. La oss kalle låtene: A B C D E F G H I J

**a) Om han skal velge 8 av de 10 låtene:**

**Hvor mange mulige kombinasjoner av 8 låter finnes det?**

```
#factorial(10)/(factorial(8)*factorial(10-8))  
choose(10, 8)
```

[1] 45

**Hvor mange mulige kombinasjoner finnes det om A, B og C skal være med?**

```
choose(7, 5)
```

```
[1] 21
```

```
#factorial(7)/(factorial(5)*factorial(7-5))
```

**b) Om han skal velge 8 av de 10 låtene:**

**Hvor mange mulige rekkefølger av 8 låter finns det?**

```
#{factorial(10)/(factorial(8)*factorial(10-8))*factorial(8)}  
choose(10, 8)*factorial(8)
```

```
[1] 1814400
```

**Hvor mange av rekkefølgene er i alfabetisk ordning?**

```
#factorial(10)/(factorial(8)*factorial(10-8))  
choose(10, 8)
```

```
[1] 45
```

Merk at i oppgave c) skal det bare velges ut 6 låter

**c) Om han skal velge 6 av de 10 låtene, men må ha med B om han har med A (om motsatt), og kan**

ikke ha med J om han har med I (og motsatt): - Hvor mange mulige kombinasjoner av 6 låter finns det?

```
ab_uten_i_og_j <- choose(6, 4)
```

```
ab_enten_ij <- choose(6, 3) * 2
```

```
uten_abij <- choose(6, 6)
```

```
i_eller_j_uten_ab <- choose(6, 5) * 2
```

```
print(ab_uten_i_og_j + ab_enten_ij + uten_abij + i_eller_j_uten_ab)
```

[1] 68

```
#(choose(6, 4) + choose(6, 3) + choose(6, 3) + choose(6, 6) + choose(6, 5) + choose(6, 5))
```

**d) Om han skal velge 8 av de 10 låtene og velger tilfeldig:**

**Hva er sannsynligheten for at A kommer med?**

```
#antall gunstie delt på antall mulige  
choose(9, 7) / choose(10, 8)
```

[1] 0.8

**Hva er sannsynligheten for at minst en av A og B kommer med?**

```
1 - (1 / choose(10, 8))
```

[1] 0.9777778

## Oppgave 4

Det er kjent at 10% av laksepopulasjonen i et vassdrag har en bestemt parasitt, her kalt parasitt A. Det er også kjent at 20% av populasjonen har en parasitt kalt parasitt B. Vi veit og at av de laksene som har parasitt A (gitt parasitt A) vil 50% ha parasitt B (betinge sannsynlighet 0.5).

**a) Er hendelsene  $A = \text{"laks har parasitt A"}$  og  $B = \text{"laks har parasitt B"}$  uavhengige?**

Dersom  $P(B|A) = P(B)$  eller  $P(A|B) = P(A)$  så er de uavhengige. Men  $P(B) = 0.2$  men  $P(B|A) = 0.5$  så de er ikke uavhengige.

**Er de disjunkte?**

Nei, fordi  $P(A \cap B) > 0$ .

**b) Regn ut sannsynligheten for at en laks har begge parasittene  $P(A \cap B)$ .**

```
b_gitt_a = 0.5
p_a = 0.1
p_b = 0.2

p_og_b = b_gitt_a * p_a
p_og_b
```

```
[1] 0.05
```

**Regn ut sannsynligheten for at en laks har minst en parasitt.**

```
p_eller_b = p_a + p_b - p_og_b
p_eller_b
```

```
[1] 0.25
```

**c) Gitt at en laks har minst en parasitt, hva er sannsynligheten for at den har parasitt A?**

```
p_a_gitt_eller_b = p_a / p_eller_b
p_a_gitt_eller_b
```

```
[1] 0.4
```

**Gitt at en laks ikke har parasitt A, hva er sannsynligheten for at den har parasitt B?**

```
p_b_gitt_kompli_a = (p_b - p_og_b) / (1 - p_a)
p_b_gitt_kompli_a
```

```
[1] 0.1666667
```

**Vi tenker oss at  $n = 10$  laks blir sjekka for parasitt A.**

**d) Hva er sannsynligheten for at minst en laks har parasitt A?**

```
1 - (1 - 0.1)^10
```

```
[1] 0.6513216
```



## Hva er sannsynligheten for at akkurat en laks har parasitt A?

```
choose(10, 1) * 0.1^1 * (1 - 0.1)^(10-1) #må dobbelsjekke når jeg har tid. eksempel 5.3
```

```
[1] 0.3874205
```

## Oppgave 5

For en type backup-batterier er det kjent at levetida (tida, i år, det tar før de svikter) er gitt ved sannsynlighetstettheten

$$f(x) = \frac{2}{5}xe^{-\frac{1}{2}x^2}, \quad x > 0$$

### a) Skisser tettheten i R.

#### Vis at dette er ei ekte fordeling.

Jeg har brukt kode jeg har funnet i et pdf dokument gitt på canvas Elvebakk (2023)

```
# dford1 = function(x) x*exp(-x) # "d" for "density".
# #Litt meir korrekt er å sette andre verdier til 0, for eksempel:
# dford1 = function(x)
#   {tetthet = vector(length=length(x))
#   for (i in 1:length(x))
#     {if (x[i]>0) tetthet[i] = x[i]*exp(-x[i])
#     else tetthet[i] = 0}
#   return(tetthet)}

dford2 = function(x) (2/5)*x*exp(-1/5*x^2)

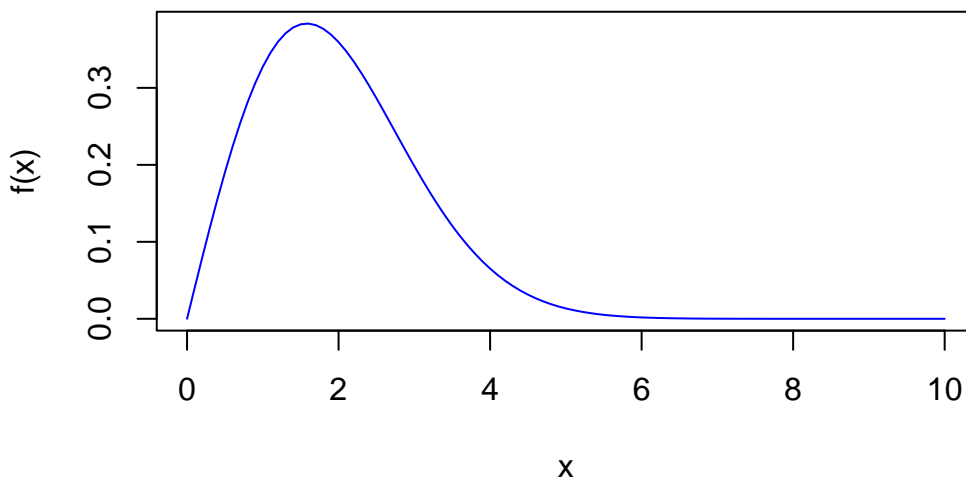
dford2 = function(x) {tetthet = vector(length=length(x))
for (i in 1:length(x))
  {if (x[i]>0) tetthet[i] = (2/5)*x[i]*exp(-1/5*x[i]^2)
  else tetthet[i] = 0}
return(tetthet)}

#curve((2/5)*x*exp(-1/5*x^2), from = 0, to = 5)

#curve(dford1,from=0,to=10,xlab="x",ylab="f(x)",main="Sannsynlighetstetthet",col="blue")

curve(dford2,from=0,to=10,xlab="x",ylab="f(x)",main="Sannsynlighetstetthet",col="blue")
```

## Sannsynlighetstetthet



```
#integrate(function(x) (2/5)*x*exp(-1/5*x^2), 0, Inf)
integrate(dford2,lower=0,upper=Inf)$value
```

```
[1] 1
```

Areal under kurven er 1, så det er en ekte fordeling.

**Finn den kumulative fordelingsfunksjonen, (  $F(x)$  ), og skissér også denne.**

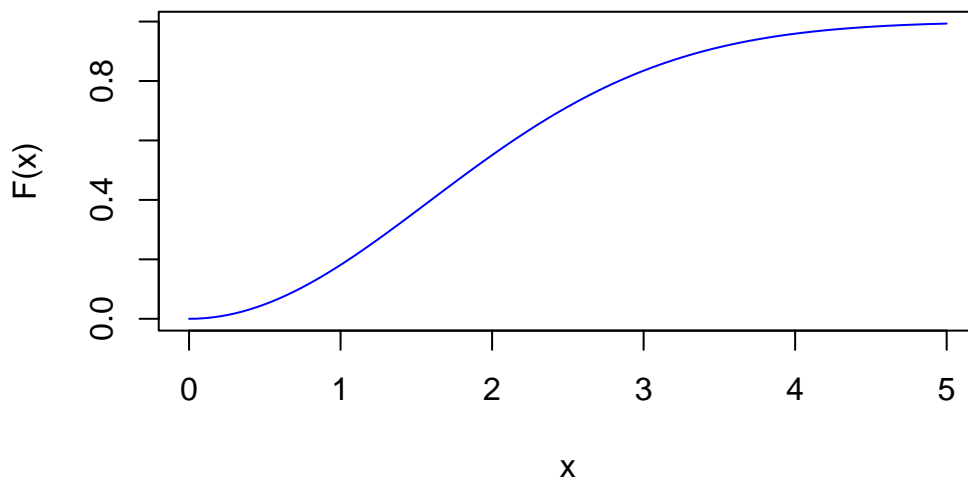
```
#curve(1 - exp(-1/5*x^2), from = 0, to = 5)
```

```
# pford1 = function(x) # "p" for "(cumulative) probablility".
# {kum = vector(length=length(x))
# for (i in 1:length(x))
# {if (x[i]>0) kum[i] = 1 - (x[i]+1)*exp(-x[i])
# else kum[i] = 0}
# return(kum)}
```

```
pford2 = function(x) # "p" for "(cumulative) probablility".
  {kum = vector(length=length(x))
  for (i in 1:length(x))
  {if (x[i]>0) kum[i] = integrate(dford2, lower=0, upper = x[i])$value
  else kum[i] = 0}
  return(kum)}
```

```
# curve(pford1,from=0,to=10,xlab="x",ylab="F(x)",main="Kumulativ fordelingsfunksjon",col="red")
curve(pford2,from=0,to=5,xlab="x",ylab="F(x)",main="Kumulativ fordelingsfunksjon",col="blue")
```

## Kumulativ fordelingsfunksjon



b) Når er det siste tidspunktet du fremdeles kan være 90% sikker på at et batteri virker?

```
#in pford2 when is y=0.9  
uniroot(function(x) pford2(c(x))[1] - 0.9, lower=0, upper=5)
```

```
$root  
[1] 3.393071
```

```
$f.root  
[1] 1.158728e-07
```

```
$iter  
[1] 9
```

```
$init.it  
[1] NA
```

```
$estim.prec  
[1] 6.103516e-05
```

3.4år

c) Finn sannsynligheten for at et tilfeldig valgt batteri skal virke etter ett år.

```
#exp(-1/5*1^2)
1- pford2(1)
```

[1] 0.8187308

Om du inspiserer et batteri etter ett år og finner at det virker, hva er sannsynligheten for at det ikke vil virke om du inspiserer det ett år seinere?

```
(pford2(2) - pford2(1))/(1 - pford2(1))
```

[1] 0.4511884

## Oppgave 6

Vi antar følgende simultane sannsynlighetstetthet for de stokastiske variablene X og Y:

$$f(x, y) = \begin{cases} 2e^{-(2x+y)}, & x > 0, y > 0 \\ 0, & \text{ellers} \end{cases}$$

a) Finn de marginale fordelingene (sannsynlighetstetthetene) for X og Y?

$$f_X(x) = 2e^{-2x}$$

$$f_Y(y) = e^{-y}$$

Integrerer for x

$$f_X(x) = \int_0^{\infty} 2e^{-(2x+y)} dy$$

$$f_X(x) = 2 \left( -e^{-(2x+y)} \right)_{y=0}^{y=\infty}$$

$$f_X(x) = 2 \left( -e^{-(2x+\infty)} + e^{-(2x+0)} \right)$$

$$f_X(x) = 2 \left( 0 - (-e^{-(2x+0)}) \right)$$

$$f_X(x) = 2e^{-2x}$$

Gjør det samme for y

$$f_Y(y) = \int_0^{\infty} 2e^{-(2x+y)} dx$$

$$f_Y(y) = 2 \left( \frac{-1}{2} e^{-(2x+y)} \right)_{x=0}^{x=\infty}$$

$$f_X(x) = 2 \left( 0 - \left( \frac{-1}{2} e^{-(0+y)} \right) \right)$$

$$f_X(x) = 2 \left( \frac{1}{2} e^{-y} \right)$$

$$f_X(x) = e^{-y}$$

```
import sympy as sp
x, y = sp.symbols('x y')

f_xy = 2*sp.exp(-(2*x + y))

f_X = sp.integrate(f_xy, (y, 0, sp.oo))

f_Y = sp.integrate(f_xy, (x, 0, sp.oo))

f_X, f_Y
```

$(2*\exp(-2*x), \exp(-y))$

### Hva blir den betingede fordelinga (sannsynlighetstettheten) for X?

Siden  $x > 0, y > 0$  så henter jeg formelen fra s119 i pensum boken.

$$f_{X|Y}(x|y) = \frac{2e^{-(2x+y)}}{e^{-y}}$$

$$f_{X|Y}(x|y) = \frac{2\cancel{e^{-(2x+y)}}}{\cancel{e^{-y}}}$$

$$f_{X|Y}(x|y) = 2e^{-2x}$$

### Er X og Y uavhengige?

Ja den betingede sannsynlighetstettheten er ikke avhengig av y

### Nå er vi interessert i hva som blir summen av X og Y.

```
#sum of x and y in  $2e^{-(2x+y)}$ 
```

**b) Finn ved integrasjon av den simultane sannsynlighetstettheten**

(  $f(x,y)$  ) hva som blir sannsynligheten for at summen av  $X$  og  $Y$  er under 2. Det vil si  
finn (  $P(X + Y < 2)$  ).

## Referanser

- Elvebakk, G. (2023). *Introduksjon til r i STA-1001: Del 2: Funksjoner og fordelingsfunksjoner*. <https://uit.instructure.com/courses/33518/files/2871351?wrap=1>
- Kristensen, Ø. & Wikan, A. (2019). *Sannsynlighetsregning og statistikk - for høyere utdanning* (2nd ed., p. 29). Fagbokforlaget.