

LESSON 3. PROBABILITY DISTRIBUTIONS

3.1. Probability

- Probability experiment, outcome and sample space
- Classical and empirical probability
- Law of large numbers

3.2. Random variables

3.3. Probability distributions and cumulative probability distributions

3.4. Types of probability distributions

- Discrete
- Continuous

3.5. Some distributions used in inferential statistics

Statistisk Dataanalyse

Lesson 3: Assignments

1. Find the sample space for the gender of the children if a family has three children. Use B for boy and G for girl.

- I tilfældet, kan det ses at vi har 3 børn og 2 køn.
- Vi har defineret det således gennem bogstaver: G = Pige og B = Dreng.

Hvordan kommer de forskellige udfald til at se ud?

Udfald	3 B	2 B / 1 G	1 B / 2 G	3 G	2 G / 1 B	1 G / 2 B	2 G / 1 B	2 B / 1 G
Samlet	BBB	BBG	BGG	GGG	GGB	GBB	GBG	BGB

2. If a family has three children, find the probability that all the children are girls.

- Sandsynlighed for at alle udfald ud af de 3 børn, ender med at være piger er svarende til $\frac{1}{8}$.

3. If the probability that a person lives in an industrialized country of the world is $\frac{1}{5}$, find the probability that a person does not live in an industrialized country.

- Sandsynligheden for, at den anden person ikke lever i et (ikke-industrialiseret) land er $\frac{4}{5}$.
- Grunden til at svaret er $\frac{4}{5}$, er fordi det er $\frac{1}{5}$ for at leve i et industrialiseret land men fordi vi har 4 sandsynlighed tilbage i tælleren, derfor ender vi med at sige at det er $\frac{4}{5}$.

4. The table below contains information on the number of daily emergency service calls received by the volunteer ambulance service of Happytown for the last 50 days: 22 days of which 2 emergency calls were received, 9 days of which 3 emergency calls were received, 8 days of which no emergency calls were received, etc.

Number of Service Calls per Day (X)	Number of Days (fj)
0	8
1	10
2	22
3	9
4	1
Total	50

What is the probability that 2 or more emergency calls are received on a day?

- Fordi vi er blevet bedt om at kigge på sandsynligheden for at 2 eller flere nødopkald er modtaget i en dag, skal vi først vende blikket imod den røde del af tabellen.

Number of Service Calls per Day (X)	Number of Days (fj)
0	8
1	10
2	22
3	9
4	1
Total	50

Statistisk Dataanalyse

- Vi kan se, at fordi vi modtager 2+3+4 flere opkald per. Dag kan vi derfor sige, at vores sandsynlighed ender med at være $9/50$ som er 0,18.

5. Determine which of the following distributions is a cumulative probability distribution.

i)

X	1	2	3	4	5
P(X≤x)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{9}{20}$	$\frac{6}{8}$	$\frac{20}{20}$

ii)

X	22	33	44	55	66	77
P(X≤x)	-0.4	0.2	0.4	0.7	0.8	1

iii)

X	0	3	5	6
P(X≤x)	0.25	0.25	0.25	0.25

iv)

X	0	2	4	6
P(X≤x)	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{4}{16}$

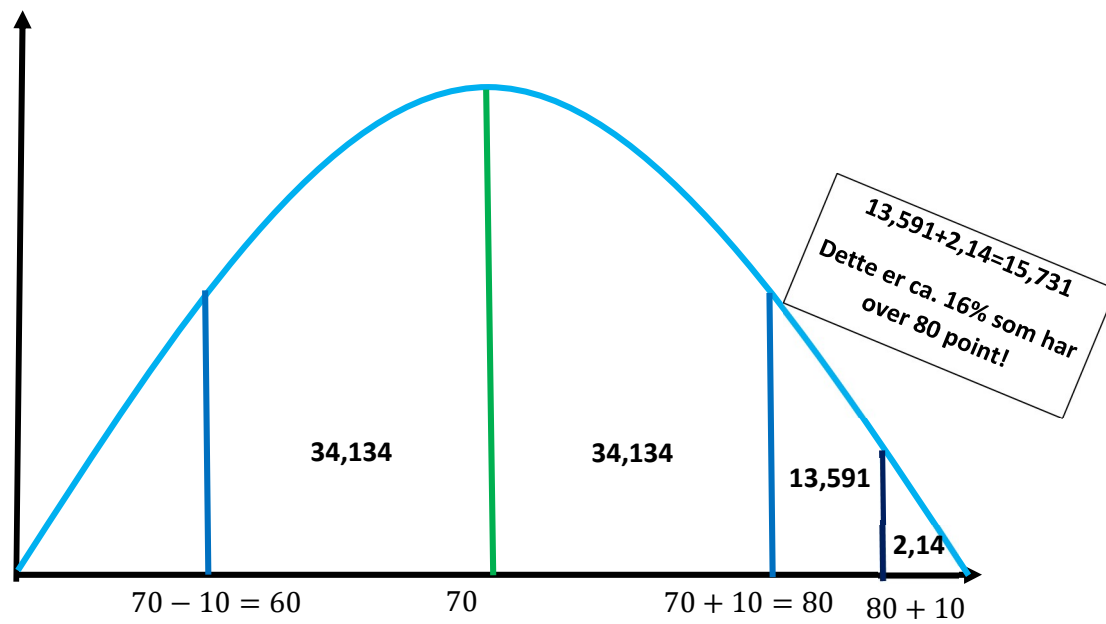
- Vi kan se at vores facit ender med at være ii.
- Grunden er tydeligt og dette er fordi vi kan se at når vi kaster en sandsynlighed flere gange, desto mere resulterer det i at vores tal bliver højere og kommer tættere på 1.

6. The grades of a group of 1000 students in an exam are normally distributed with a mean of 70 and a standard deviation of 10. Approximately, how many students have grades greater than 80?

- a) 680 students
- b) 840 students
- c) 160 students
- d) 50 students
- e) 320 students

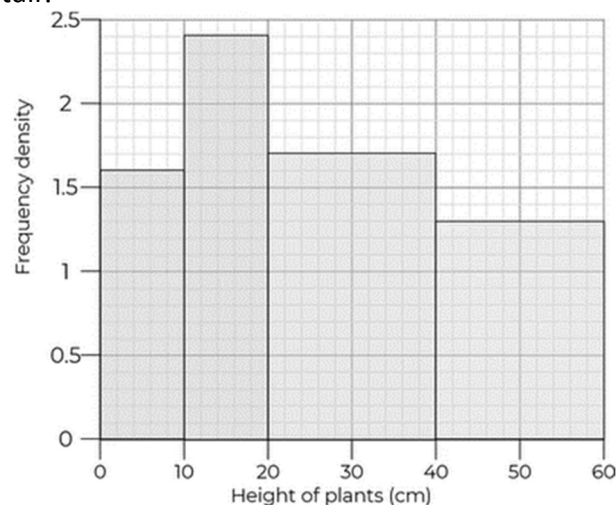
- Husk, altid at når der ikke bliver nævnt noget omkring konfidensniveauet, så er det svarende til $95\%=0,05$.
- Dette betyder, at når vi lægger vores middelværdi sammen med vores standardafvigelse, skal vi tage højde for at multiplicere med 2.
- Jeg har lavet et billede her for at give et overblik.

Statistisk Dataanalyse



- Vi kan se, at hvis man lægger ekstra 10 til ved 80, så kan det ses at vi får 90.
- Intervallet mellem middelværdien og konfidensintervallet på begge sider af middelværdien svarende til 68%.
- Men hvis man vender blikket over mod 80 og frem ad, så kan det ses at det er $13,591 + 2,14 = 15,731$ hvilket betyder at det svarende til ca. 16%.
- Fordi det kan vi sige, at 16% ud af de 1000 har scoret over 80 point i karakteren.
- Derfor bliver svaret lig med 160 elever som har over 80 point.

7. We visited a field with plants. We measured the height of each plant and built the histogram below. We then chose one plant at random. What is the probability that the plant is under 30 cm tall?



- Vi kan se, at langs x-aksen findes der parametre for højden af planterne i cm.

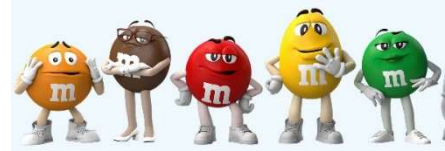
Statistisk Dataanalyse

- Fordi vi bliver spurgt om højden under 30 cm, så er det fra 0 til 20.
- Derfor kan vi sige at 3 højder ud af de hele 7 totale er vores sandsynlighed = $3/7$.

8. M&M sweets are of varying colors and the different colors occur in different proportions. The table below gives the probability that a randomly chosen M&M has each colour, but the value for orange candies is missing.

Color	Brown	Red	Yellow	Green	Orange
Probability	0.2	0.3	0.2	0.1	?

You draw an M&M at random from a packet. What is the probability that you get either a green one or an orange one?

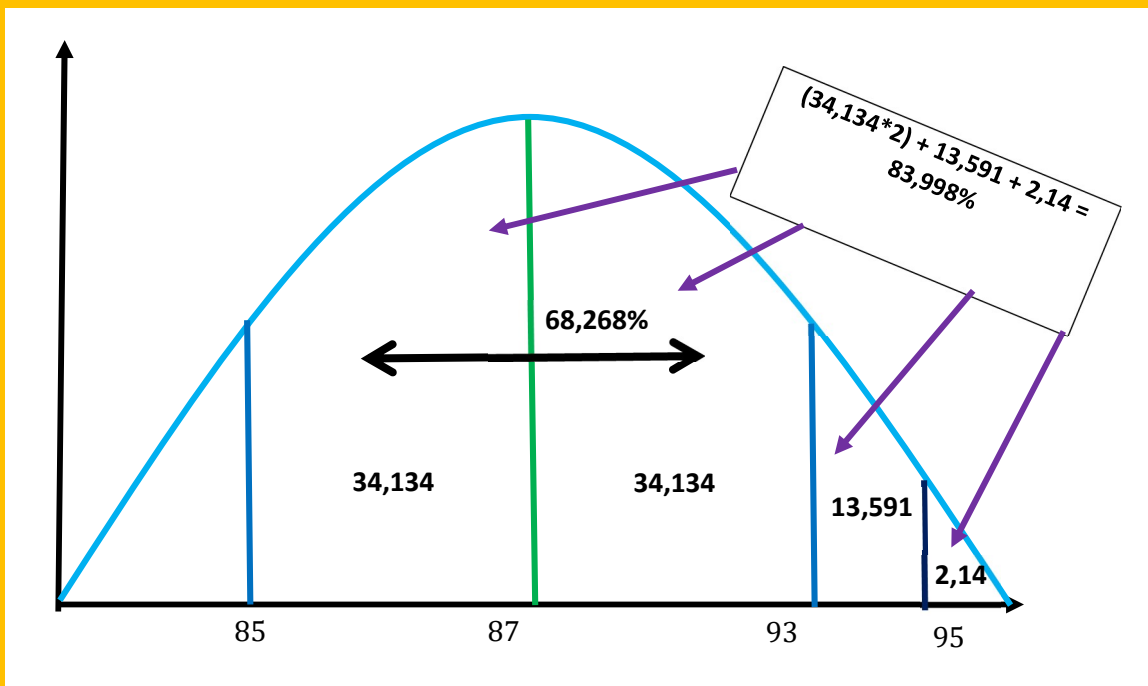


- Fordi vi bliver spurgt om sandsynlighed om det ene eller det andet, så kan vi sige at vores svar bliver $1/5$.
- Grunden til at det bliver $1/5$, er fordi vi har (enten) mulighed for at vælge det ene eller det andet. Derfor bliver det $1/5$.

9. What percent of cases are likely to be between 85 and 93 in a normal distribution with mean 87 and variance 4?

- a. 83.85%
- b. 30.72%
- c. 49.87%
- d. 69.02%
- e. none of these

- Vi kan se, at vores svar bliver 83,85%.
- Grunden er meget tydeligt og dette er fordi vi kan se at vores konfidensinterval mellem 85 og 93 er svarende til 68%.



Statistisk Dataanalyse

- Hvis man kigger yderligere nærmere på billedet, kan det ses at vi ønsker at gå længere fremad end de 93 i intervallet. Vi kan også se, at vi har yderligere 13,591% og mere yderligere 2,14% at beskæftige os med.
- Derfor kan $13,591 + 68 + 2,14$ lægges sammen, som er 83,731%

10. A survey found that one out of five Americans say he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

- Vi kan se, at der er 10 personer som er udvalgt tilfældigt og derfor kan vi se at der er snak om personer fra to køn.
- Fordi vi har disse muligheder, kan vi sige at der sandsynlighed for $2/10$.

11. Suppose a loaded die has the following model:

Face	1	2	3	4	5	6
Probability	0.3	0.1	0.1	0.1	0.1	0.3

If this die is thrown and the top face shows an odd number,

- What is the probability that the die shows a four?
- What is the probability that the die shows a 1?

A. Vi starter allerførst med at tælle hvor mange lige tal vi har i terningekastet

- Vi kan i vores tilfælde se, at vi har 3 lige tal i vores sandsynlighed, ud af de 6 andre muligheder. Derfor kan vi sige, at sandsynligheden er $3/6$ og som er den samme for at lande i ulige tal i rækken.

B. Først skal vi tælle, hvor mange sandsynligheder der er for at lande på 0,1.

- Vi kan i vores tilfælde se, at spørgsmålet er formuleret meget mærkeligt, men de vil i virkeligheden gerne have dig til at kigge på en underkategori under 4 som er ved probability rækken.
- Dette betyder, at fordi vi har 4 sandsynligheder med 0,1 ud af de 6, hvor en af dem er ved face 4. Derfor kan vi sige, at vores svar ender med at være $4/6 = 0,66$.