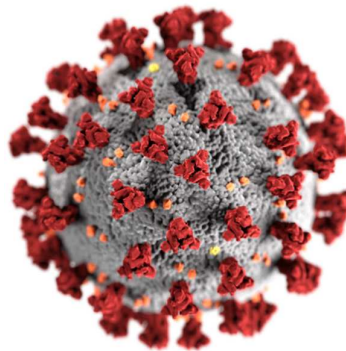


Corona-pandemien i 2020

I december 2019 brød en ny sygdom ud i Wuhan i Kina. Sygdommen fik navnet Covid-19 (Coronavirus disease 2019), og den skyldes en virus, som fik navnet SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2). Man mener, at sygdommen spredte sig med rejsende ud fra Wuhan og ud fra Kina til de fleste lande i verden, og den 11. marts 2020 blev Covid-19 erklæret en pandemi, altså en verdens-omspændende epidemi. Samme aften valgte den danske regering at lukke Danmark ned for at begrænse smittespredningen.



Umiddelbart efter udbruddet af coronavirus indledte forskere fra hele verden et kapløb for at forstå virussen, hvordan den smitter, hvor farlig den er og hvorfor nogle patienter er mere udsatte og dør af sygdommen, mens andre gennemgår sygdommen stort set uden symptomer. Sideløbende har man forsøgt at udvikle effektive vacciner uden alvorlige bivirkninger.

Der findes allerede store mængder forskningsbaserede data, der kan analyseres statistisk, for at man kan blive klogere på Covid-19.

Bemærk følgende:

- Decimaltal i opgaverne er angivet med engelsk decimalseparator (.)
 - For de opgaver, der har større mængder data tilknyttet, kan data downloades fra Digital Eksamen portalen i csv-format (data separeret med semikolon, ';'). Opgavens nummer fremgår af filen titel.
 - Nogle delopgaver benytter resultatet fra en tidligere delopgave. Hvis du ikke kunne løse den, kan du blot antage en realistisk værdi for resultatet og regne videre med det.
 - Selv om statistikpakken R er brugt i undervisningen, er det tilladt at bruge andre værktøjer. I nogle delopgaver kan det være enklere at bruge f.eks. Excel.
-

Opgave 1 – Byen med det første dødsfald i Italien

Ifølge avisen The Guardian blev det første Covid-19 relaterede dødsfald i Italien registreret i byen Vo' vest for Venedig. Patienten var Adriano Trevisan, en 78-årig mand, som døde den 22. februar 2020. Den 6. marts besluttede myndighederne at lukke for adgangen til Vo' og at teste samtlige 3300 indbyggere, uanset om de havde symptomer eller ej. Det viste sig, at 89 indbyggere testede positiv og blev sat i isolation. 14 dage senere, den 20. marts, blev alle indbyggerne testet igen. Nu blev kun 6 testet positiv, og alle 6 var blandt de 89, der allerede var sat i isolation. 14 dage senere, den 3. april, var sygdommen udryddet i byen (ingen blev testet positiv).

Lad C_1 , C_2 og C_3 angive hændelserne, at man som borger i Vo' blev testet positiv for Covid-19, henholdsvis den 6. marts, den 20. marts og den 3. april. Beregn de følgende sandsynligheder på baggrund af de ovenstående oplysninger:

- $P(C_1)$, $P(C_2)$ og $P(C_3)$
- $P(C_2 \cap C_1)$ og $P(C_2 | C_1)$

Forskerne bag undersøgelsen opdagede det skræmmende resultat, at ud af de 89 indbyggere, der oprindeligt blev testet positive, var der 63, der ikke udviste nogen af de karakteristiske symptomer på Covid-19. Man kan altså være smittet uden at vide det. Forskerne konstaterede også, at da de testede samtlige indbyggere den 6. marts, var der 127 indbyggere, der havde Covid-19 symptomer, selv om de var testet negativ. Lad S_1 betegne hændelsen, at man udviser Covid-19 symptomer den 6. marts og \bar{S}_1 , at man ikke gør (komplementærhændelsen).

- Hvad er $P(S_1 | C_1)$, $P(\bar{S}_1 | C_1)$ og $P(S_1 | \bar{C}_1)$?
- Hvor stor en andel af Vo' havde Covid-19 symptomer den 6. marts? M.a.o. hvad er $P(S_1)$?
- Hvis en indbygger i Vo' har Covid-19 symptomer, hvad er så sandsynligheden for, at vedkommende testes positiv? M.a.o. hvad er $P(C_1 | S_1)$?

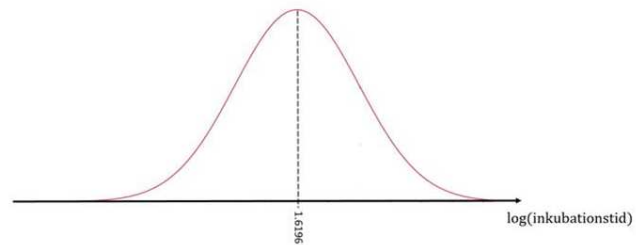
Kilde: <https://www.theguardian.com/commentisfree/2020/mar/20/eradicated-coronavirus-mass-testing-covid-19-italy-vo>

Opgave 2 – Inkubationstid

Når en ny sygdom dukker op, er det vigtigt at have et godt estimat for sygdommens inkubationstid. Inkubationstiden er det tidsrum, der går, fra en person bliver smittet, og til de første symptomer viser sig. I begyndelsen af Covid-19 epidemien blev personer, der havde været i kontakt med smittede, anbefalet at gå i karantæne i 14 dage. Man gættede på, at inkubationstiden var under 14 dage, så det var en tilstrækkeligt lang karantænetid til at opdage, om personen var smittet.

En forskergruppe fra Johns Hopkins Bloomberg School of Public Health i USA har undersøgt inkubationstiden for Covid-19. Studiet blev foretaget på 181 patienter, primært kinesere og turister, der havde været i Wuhan, og resultatet blev publiceret den 10. marts 2020 i det videnskabelige tidsskrift *Annals of Internal Medicine*.

Forskergruppen anslog først inkubationstiden for hver af de 181 patienter ud fra samtaler med patienterne. Så analyserede de disse data statistisk og fandt ud af, at inkubationstiden kan beskrives som en såkaldt 'log-normal-fordeling'. Med andre ord, hvis de tog den naturlige logaritme til inkubationstiderne, så var disse data normalfordelt med middelværdi $\mu = 1.6198$ og standardafvigelse $\sigma = 0.4387$.



R koden herunder er et eksempel, der illustrerer, hvordan forskergruppens resultater kan bruges til at estimere, hvor stor en andel af Covid-19 smittede, der har en inkubationstid på 7 døgn eller herunder (du kan evt. kopiere koden over i R):

```
## Hvor stor en andel af smittede vil have en inkubationstid på under 7 døgn?  
inku_7 = 7 # Inkubationstid, der ønskes undersøgt  
log_inku_7 = log(inku_7) # Bemærk at i R er log() den naturlige logaritme.  
# Variablen log_inku_7 = 1.9459 er den naturlige logaritme til inkubationstiden.  
  
# Vi får at vide, at den naturlige logaritme til inkubationstiden følger en  
# normalfordeling med mu = 1.6198 og standardafvigelse sigma = 0.4387  
mu = 1.6198 # Middelværdi  
sigma = 0.4387 # Standardafvigelse  
  
andel_inku_under_7 = pnorm(log_inku_7, mu, sigma) # andel_inku_under_7 = 0.7714  
# 77 pct af smittede vil have udvist Covid-19 symptomer efter en uge.
```

Beregn følgende på baggrund af forskergruppens resultater:

- Hvor stor en andel af smittede kan forventes først at udvise symptomer efter 14 døgn, altså efter karantænenes udløb?
- Forskergruppen beregnede medianen for inkubationstiden til 5.1 døgn. M.a.o. vil halvdelen af Covid-19 smittede udvikle symptomer inden der er gået 5.1 døgn. Udregn medianen for inkubationstiden og oplys resultatet.
- Hvornår har 95% af alle smittede udvist symptomer?
- Bestem det empiriske interval for den normalfordelte stokastiske variabel og omregn til et interval, som inkubationstiden ligger indenfor for næsten alle patienter.

Opgave 3 – Udvikling af antal døde i Danmark og Sverige

De danske og de svenske myndigheder har valgt meget forskellige strategier for bekæmpelsen af coronavirus. Sverige benytter en langt mindre restriktiv tilgang end Danmark og de fleste andre europæiske lande. F.eks. har Sverige undladt at lukke skoler og virksomheder. Det er interessant at undersøge, om de forskellige strategier har haft en effekt på udviklingen af smitte.

Herunder vises det samlede antal personer, der er registreret døde med Covid-19, per dag i Danmark (DK) og Sverige (SE). Dag nr. 71 i 2020 svarer til onsdag den 11. marts, hvor Skandinaviens første dødsfald blev rapporteret i nærheden af Stockholm.

		Antal døde				Antal døde				Antal døde	
Dato	Dag	DK	SE	Dato	Dag	DK	SE	Dato	Dag	DK	SE
10-03	70	0	0	23-03	83	24	25	05-04	96	179	401
11-03	71	0	1	24-03	84	32	36	06-04	97	187	477
12-03	72	0	1	25-03	85	34	42	07-04	98	203	591
13-03	73	0	1	26-03	86	41	66	08-04	99	218	687
14-03	74	0	1	27-03	87	52	92	09-04	100	237	793
15-03	75	1	3	28-03	88	65	102	10-04	101	247	870
16-03	76	1	7	29-03	89	72	110	11-04	102	260	887
17-03	77	4	8	30-03	90	77	146	12-04	103	273	899
18-03	78	4	10	31-03	91	90	180	13-04	104	285	919
19-03	79	6	10	01-04	92	104	239	14-04	105	299	1033
20-03	80	9	16	02-04	93	123	282	15-04	106	309	1203
21-03	81	13	20	03-04	94	139	333	16-04	107	321	1335
22-03	82	13	21	04-04	95	161	373	17-04	108	332	1507

Kilde: https://covid.ourworldindata.org/data/ecdc/total_deaths.csv

- For at gøre data mere sammenlignelige justeres de til antal dødsfald pr. million indbyggere. Sverige har 10.3 millioner indbyggere, Danmark har 5.8 millioner. Smid desuden data bort for dagene før dagnummer 80 pga. små talværdier. Lav en graf, der viser antal døde pr. million indbyggere i Danmark og Sverige som funktion af dag efter dagnummer 80.
- Lav på baggrund af de justerede data fra delspørgsmål a. en lineær model for hvert af de to lande:
$$C(t) = At + B$$
 hvor $C(t)$ er antal døde med Covid-19 pr. million indbyggere på dag nummer t og A og B er konstanter. Skriv modelligningen op for hvert land.
- Er de to lineære modeller gode til at beskrive data?
- På hvilket dagnummer kan hhv. Danmark og Sverige forvente at overskride 1000 døde pr. million indbyggere ifølge modellerne?
- Ofte antager man at man kan modellere den tidlige fase af en smittefarlig sygdom som en eksponentiel vækst. Fit de justerede data fra delspørgsmål a. til en eksponentiel model for hvert af de to lande:
$$C(t) = Be^{At} \Leftrightarrow \log(C(t)) = \log(B) + At$$
 hvor $C(t)$ er antal døde med Covid-19 pr. million indbyggere på dag nummer t og A og B er konstanter. Skriv modelligningen op for hvert land.
- Hvilken af de to typer model (lineær og eksponentiel) mener du, der passer bedst for hhv. Danmark og Sverige? Begrund svaret.