Exercise 1 Calc. : ✓

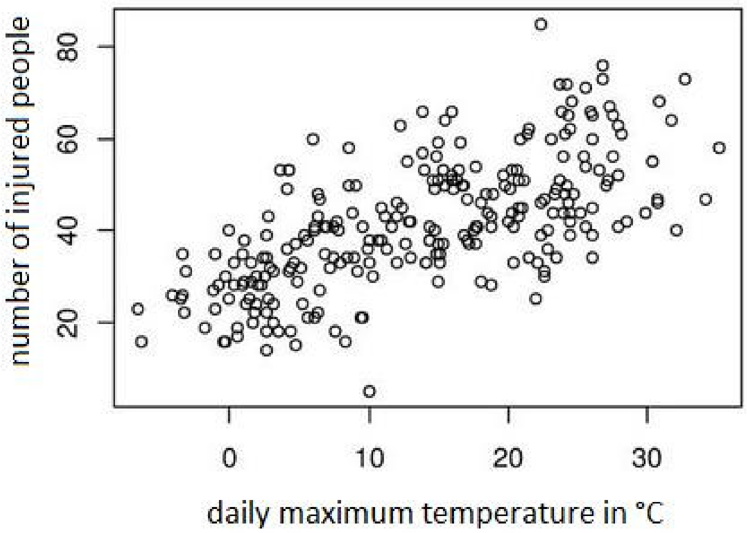
Vuoden 2002 keskimääräiset kuukausilämpötilat Luxembourgissa kirjattiin ylös. Tammikuu 2002 oli kylmin kuukausi, jolloin keskimääräinen lämpötila oli $1,6$°C. Kesäkuu 2002 oli puolestaan kuumin kuukausi, ja sen keskilämpötila oli $18,6$°C.	
1. Perustele, miksi Euroopassa keskimääräiset kuukausilämpötilot muutaman peräkkäisen vuoden ajalta noudattavat jaksollista mallia.	2 marks
2. Määritä vuoden 2002 havaintojen pohjalta tehdyn jaksollisen mallin amplitudi ja jakso.	2 marks
3. Määritä parametrit a, b, c ja d jos jaksollinen malli on muotoa: $$T(x) = a \cdot \sin(b \cdot (x - c)) + d$$ missä T on keskimääräinen lämpötila ja x kuukausi, siten että $x = 1$ vastaa tammikuuta 2002.	5 marks
Tiettynä päivänä maaliskuussa 2002 sademäärä mitattiin. Sademäärää kyseisenä päivänä voidaan mallintaa funktiolla: $$R(t) = 0,002t^3 - 0,064t^2 + 0,512t, \qquad 0 \leq t \leq 24$$ missä $R(t)$ on sademäärän nopeus (mm/h) ja t on aika tunteita.	
4. Kuvaile lyhyesti tämän päivän sademääriä eri aikoina. Kerro, milloin sataa eniten ja milloin vähiten.	3 marks
Tyhjä lasisylinteri laitettiin ulos tämän päivän aikana auttamaan näkemään, kuinka paljon sadetta oli satanut.	
5. Piirrä kuvaaja, joka esittää veden korkeutta lasissa, joka asetettiin tyhjänä ulos.	3 marks
6. Laske kokonaissademäärä päivän aikana.	2 marks

Vuonna 2002 Luxemburgissa oli 195 sadepäivää ja 170 sateetonta päivää. Voidaan olettaa, että jokaisen päivän kohdalla sateen todennäköisyys pysyy samana. Vuotta myöhemmin meteorologit haluavat tutkia, oliko vuosi 2003 sateisempi. Valitettavasti tietoa oli hävinnyt, ja heillä oli kerättynä tietoa vain 30 peräkkäisestä päivästä.

7. Laske, millä todennäköisyydellä päivä oli sadepäivä vuonna 2002, jos oletetaan, että sadepäivät jakaantuivat tasaisesti koko vuodelle. — 1 mark

8. Käytä NHST-menetelmää, kuinka monena päivänä pitää sataa vuonna 2003, jotta voitaisiin sanoa, että vuosi 2003 oli sateisempi kuin vuosi 2002. Käytä 5% merkitsevyystasoa. Oletetaan, että sateen todennäköisyys päivää kohden pysyy samana joka vuosi. — 5 marks

Alla olevassa diagrammissa on esitetty maksimilämpötila ja liikenneonnettomuuksissa vahingoittuneiden ihmisten määrä Berliinissä pitkällä aikavälillä.



9. **Kuvaile,** millainen korrelaatio näiden kahden muuttuja välillä on. — 1 mark

10. **Selitä,** miksi loukkaantuneiden ihmisten määrä voisi riippua maksimilämpötilasta. — 1 mark

Exercise 2 Calc. : ✓

Covid-testausasemalla testattiin eräänä päivänä 19 oireista ihmistä, ja 6 heistä sai positiivisen testituloksen. Samana päivänä testattiin 87 ihmistä, joilla ei ollut oireita, ja 85 heistä sai negatiivisen testituloksen.

1. Näytä, että positiivisen testituloksen saaminen riippuu siitä, onko ihmisellä oireita tai ei. — 2 marks

Suojellakseen henkilötietoja testitikut merkitään koodilla, jossa on kaksi kirjainta (A–Z eli 26 vaihtoehtoa) ja neljä numeroa (0–9). Samaa kirjainta tai numeroa voidaan käyttää koodissa useamman kerran.

2. Laske, kuinka monta erilaista koodiyhdistelmää on mahdollista tehdä. — 2 marks

Muutamien kuukausien kuluttua saadaan selville, että ihmisistä, joilla ei ole oireita, 1,7% saa testissä positiivisen tuloksen. Eräs yritys, jossa on 20 työntekijää (kaikki oireettomia) testauttaa kaikki työntekijänsä.

3. Kerro kaksi oletusta, jotka pitää tehdä, jotta tätä tilannetta voidaan mallintaa binomijakaumalla. — 2 marks
4. Laske, millä todennäköisyydellä ainakin 1 työntekijä saa positiivisen testituloksen. — 3 marks

Myös eräs toinen yritys testauttaa kaikki työntekijänsä. Oletetaan, että tilannetta voidaan mallintaa binomijakaumalla:

$$B(84; 0,02; k) = \binom{84}{k} \cdot 0,02^k \cdot 0,98^{84-k}.$$

5. Mitä luvut 84, $0,02$ ja $0,98$ kuvaavat tässä tilanteessa? — 3 marks

5. maaliskuuta 2020 Italiasta kotiin palaava mies oli ensimmäinen COVID-19 -positiiviseksi testattu henkilö Luxemburgissa. Tätä päivää merkitään päivänä 0 alla olevassa taulukossa. Siinä on esitetty sairastuneiden määrä seuraavina päivinä ensimmäisestä tapauksesta.

Jour	0	1	2	3	4	5	6
Nombre	1	3	4	5	5	7	7

6. **Piirrä** sirontakuvaaja näistä arvoista, ja lisäksi niihin sovitettu lineaarinen ja eksponentiaalinen regressiomalli. — 3 marks
7. **Määritä** f-kohdan mallien (lineaarisen ja eksponentiaalisen) yhtälöt. — 2 marks
8. **Selitä**, miksi alkuvaiheessa oli vaikea päättää, kumpi malli (lineaarinen vai eksponentiaalinen) sopii paremmin kuvaamaan viruksen leviämistä. — 2 marks

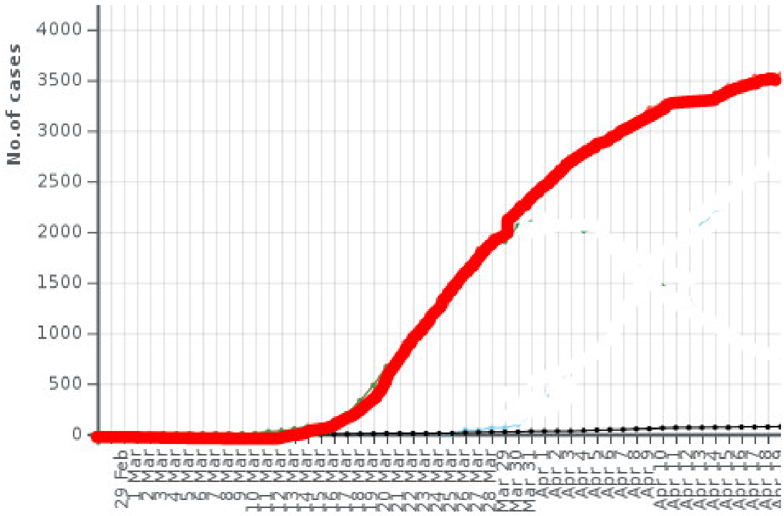
Seitsemän päivän jälkeen tehtiin uudet mallit kuvaamaan viruksen leviämistä, jotta sitä voitaisiin ennustaa tarkemmin. Näissä malleissa t tarkoittaa aikaa päivinä (viruksen saapumisesta maahan):

$$A(t) = 1,35567 \cdot 1,46977^t \qquad B(t) = 12,4396 \cdot t - 34,8571$$

Päivänä 16 COVID-19 sairastuneita rekisteröitiin 670 Luxemburgissa.

9. **Laske**, kuinka paljon sairastuneita on mallien A ja B mukaan Luxemburgissa ja vertaa mallien antamaa tulosta oikeaan arvoon. Päätä, kumpi malleista sopii paremmin tilanteeseen ja perustele vastauksesi. — 2 marks

Alla olevassa kuvaajassa on esitetty rekisteröityjen sairastumisten määrä Luxemburgissa ensimmäisen 4 viikon ajalta.



10. **Anna** kaksi mahdollista syytä siihen, miksi kasvu vähenee voimakkaasti myöhemmässä vaiheessa. — 2 marks

Tätä kuvaajaa voidaan mallintaa funktiolla:

$$C(t) = \frac{3404}{1 + 193 \cdot e^{-0,233 \cdot t}}$$

11. **Määritä**, minä päivänä tartuntojen määrän muutos oli suurin. — 2 marks