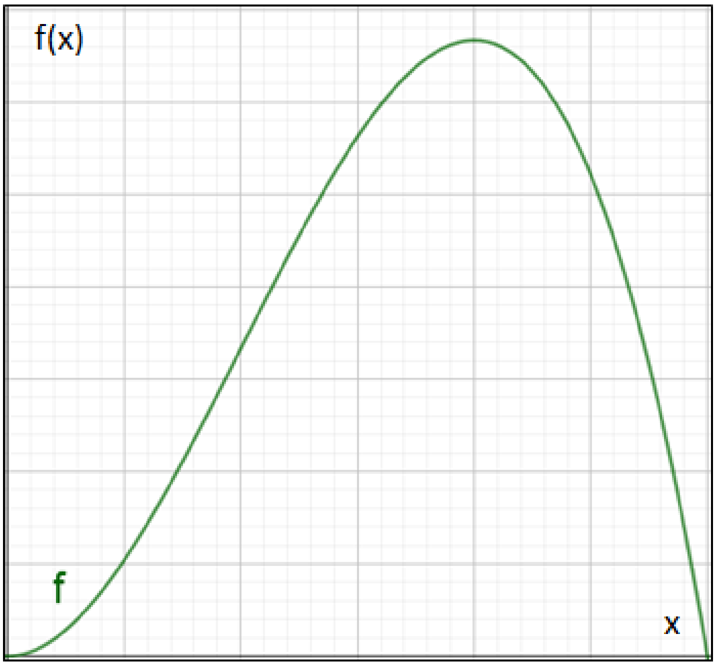
# S7 3P – BAC2023 mallikoe 2

## A-osa (ilman laskinta)

**A1.** Leikkikentällä olevaa mäkeä voidaan mallintaa funktiolla

$f(x) = -\frac{1}{3}x^3 + x^2$, kun $x > 0$ missä $x$ on vaakasuora etäisyys metreinä ja $f(x)$ on korkeus metreinä. Funktion kuvaaja on esitetty alla. Määritä mäen korkeus.




**A2.** Ruokalan kokin mukaan 20 % kaikista 2500 oppilaasta ei ole tyytyväisiä ruoan laatuun. Oppilaskunnan mielestä osuus on suurempi kuin 20 %. Oppilaskunta päättää kysyä 40 oppilaalta heidän mielipidettään.

**a)** Selitä, pitäisikö tässä hypoteesin testauksessa käyttää vasemman- vai oikeanpuoleista testiä. Perustele vastauksesi.

**b)** Mitä nollahypoteesia pitäisi käyttää tähän NHST-testiin? Entä mitä vastahypoteesia?

**c)** Määritä kriittinen arvo *k* oheisen taulukon avulla (missä *k* on tyytymättömien opiskelijoiden määrä), jos riskitasoksi asetetaan 5 %. Tulkitse, mitä tämä arvo tarkoittaa.

| $k$ | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|
| $P(X \geq k)$ | 0.563 | 0.407 | 0.268 | 0.161 | 0.088 | 0.043 | 0.019 | 0.008 |

**A3.** Pieni supermarketketju työllistää 900 ihmistä, ja vain 10 heistä työskentelee johdossa, mutta vain yksi johtajista on nainen. Muut 809 naista ovat töissä kaupoissa työntekijöinä. Näytä, että naisena ja johtoasemassa oleminen eivät ole toisistaan riippumattomia tapahtumia.

**A4.** Aviopari tarvitsee negatiivisen Covid-testin, jotta he voivat vierailla ystäviensä luona ulkomailla. Tiedetään, että 20 % testeistä antaa väärän negatiivisen tuloksen, kun taas väärän positiivisen tuloksen todennäköisyys on lähellä nollaa. Oletetaan, että jos toinen parista on saanut tartunnan, niin myös toinen on saanut sen.

Selitä, miksi tässä on kyse Bernoulli-kokeesta, ja että, jos molemmat tekevät testin, väärän negatiivisen tuloksen suuruus on enää 4 %.

**A5.** Alla olevassa kuvassa on esitetty uusien Covid-19 tapausten määrä Saksassa 3 viikon ajalta lokakuussa 2020. Tapausten määrää tulevaisuudessa voidaan ennustaa yhdistämällä kaksi matemaattista mallia.


Kerro, mitkä nämä kaksi matemaattista mallia voisivat olla ja perustele vastauksesi. Ennusta, minä päivänä tulevaisuudessa seuraava maksimi voisi olla, jos tapausten määrä noudattaa samaa mallia kuin kuvassa esitetyn kolmen viikon aikana.

**A6.** Laboratoriossa tutkitaan petrimaljassa olevien bakteerien määrää. Käy ilmi, että tietyissä olosuhteissa bakteerien määrää voidaan mallintaa funktiolla:

$$N(t) = 10\ 000 \cdot e^{\ln(1{,}03) \cdot t},$$

missä $N(t)$ on bakteerien määrä ja $t$ aika päivissä.

a. **Laske** bakteerien määrä alussa sekä laske, kuinka monta prosenttia bakteerien määrä kasvaa päivän aikana.
b. **Laske** bakteerien määrä yhden päivän jälkeen.
c. **Selitä**, miksi tätä mallia ei voi käyttää, kun t kasvaa hyvin suureksi.

**A7.** Ovatko seuraavat väittämät totta vai tarua? Perustele vastauksesi. Huomaa, että saat pisteitä vain, jos sekä vastaus että perustelu ovat oikein.

**a)** Jos lämpötila *T(x)* aidosti kasvava, niin silloin *T'(x) < 0*.

**b)** Kaikki jaksolliset ilmiöt voidaan mallintaa sinifunktiolla.

**c)** Kolmesta oppilaasta voidaan muodostaa 9 erilaista jonoa.

**d)** Kun noppaa heitetään kerran, odotusarvo on 3,5.

**e)** Kun 10 ihmistä valitaan suuresta joukosta, naisten määrää voidaan mallintaa binomijakaumalla.

**A8.** Päivän pituutta tunteina *L(t)* tutkittiin eräässä paikassa vuoden ajan. Sitä voidaan mallintaa funktiolla:

$$L(t) = 4 \cdot \sin\left(\frac{2\pi}{365}x\right) + 12,$$

missä *t* on aika päivinä vuoden alusta.

Kerro, mitä saadaan laskemalla $\int_0^{365} L(t)dt$ ja miksi tulos on sama kuin $12 \cdot 365 = 4380$.

**A9.** Binomi- ja normaalijakauma ovat kaksi paljon käytettä todennäköisyysjakaumaa.

**a)** Kerro yksi ominaisuus, millä tavoin nämä kaksi jakaumaa eroavat toisistaan, ja yksi ominaisuus, joka niillä on yhteistä.

**b)** Kerro esimerkki parametrista, jota tarvitaan normaalijakaumaa määritettäessä.

Normaalijakaumaa voidaan kuvata funktiolla:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

**c)** Selitä, miksi

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

**A10.** Uusi kone tunnistaa verestä, jos urheilija on käyttänyt dopingia. Nimetään tapahtumat:

P: Doping-testi on positiivinen

D: Urheilija on käyttänyt dopingia

Konetta testattiin, ja havaittiin, että 100 dopingia sisältävästä verestä kone tunnistaa 90. Se antaa myös väärän positiivisen tuloksen 5 % todennäköisyydellä. Voidaan olettaa, että joka 10. urheilija on keskimäärin käyttänyt dopingia.

Lasketaan todennäköisyys, että urheilija on käyttänyt dopingia, kun testi on positiivinen:

a) Esitä kaikki tehtävää varten tarvittavat tiedot matemaattisesti oikein.

b) Laske todennäköisyys, että urheilija on käyttänyt dopingia, kun testitulos on positiivinen.

<u>B-osa (laskimen kanssa)</u>

**B1.**

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.

**a)** Perustele, miksi Euroopassa keskimääräiset kuukausilämpötilot muutaman peräkkäisen vuoden ajalta noudattavat jaksollista mallia.

**b)** Määritä vuoden 2002 havaintojen pohjalta tehdyn jaksollisen mallin amplitudi ja jakso.

**c)** Määritä parametrit a, b, c ja d jos jaksollinen malli on muotoa:

$T(x) = a \cdot \sin(b(x-c))+d$

missä T on keskimääräinen lämpötila ja x kuukausi, siten että x = 1 vastaa tammikuuta 2002.

Tiettynä päivänä maaliskuussa 2002 sademäärä mitattiin. Sademäärää kyseisenä päivänä voidaan mallintaa funktiolla:

$R(t) = 0{,}002t^3 – 0{,}064t^2 + 0{,}512t , \ 0 \leq t \leq 24$

missä R(t) on sademäärän nopeus (mm/h) ja t on aika tunteita.

**d)** Kuvaile lyhyesti tämän päivän sademääriä eri aikoina. Kerro, milloin sataa eniten ja milloin vähiten.

**e)** Piirrä kuvaaja, joka esittää veden korkeutta lasissa, joka asetettiin tyhjänä ulos.

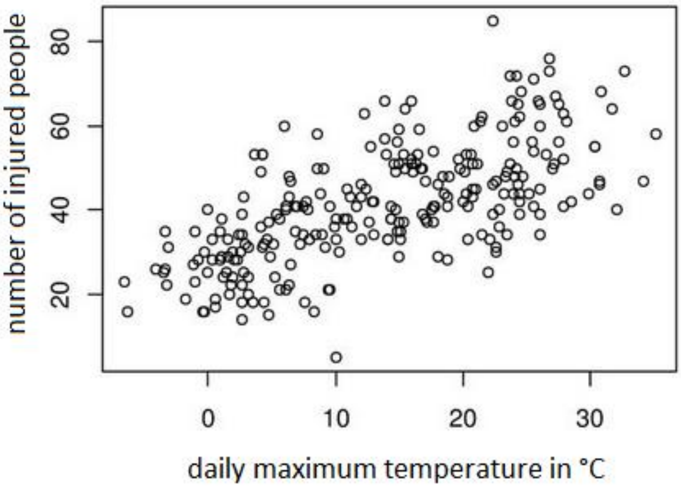
**f)** Laske kokonaissademäärä päivän aikana.

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.

**g)** Laske, millä todennäköisyydellä päivä oli sadepäivä vuonna 2002, jos oletetaan, että sadepäivät jakaantuivat tasaisesti koko vuodelle.

**h)** 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.

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




**i) Kuvaile,** millainen korrelaatio näiden kahden muuttuja välillä on.

**j) Selitä,** miksi loukkaantuneiden ihmisten määrä voisi riippua maksimilämpötilasta.

**B2.**

**<u>OSA I</u>**

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.

**a)** Näytä, että positiivisen testituloksen saaminen riippuu siitä, onko ihmisellä oireita tai ei.

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.

**b)** Laske, kuinka monta erilaista koodiyhdistelmää on mahdollista tehdä.

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.

**c)** Kerro kaksi oletusta, jotka pitää tehdä, jotta tätä tilannetta voidaan mallintaa binomijakaumalla.

**d)** Laske, millä todennäköisyydellä ainakin 1 työntekijä saa positiivisen testituloksen.

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

$B(84; 0{,}02; k) = \binom{84}{k} \cdot 0.02^k \cdot 0.98^{84-k}$.

**e)** Mitä luvut 84, 0,02 ja 0,98 kuvaavat tässä tilanteessa?

**OSA II**

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.

| Day | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Number | 1 | 3 | 4 | 5 | 5 | 7 | 7 |

**f) Piirrä** sirontakuvaaja näistä arvoista, ja lisäksi niihin sovitettu lineaarinen ja eksponentiaalinen regressiomalli.

**g) Määritä** f-kohdan mallien (lineaarisen ja eksponentiaalisen) yhtälöt.

**h) Selitä**, miksi alkuvaiheessa oli vaikea päättää, kumpi malli (lineaarinen vai eksponentiaalinen) sopii paremmin kuvaamaan viruksen leviämistä.

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

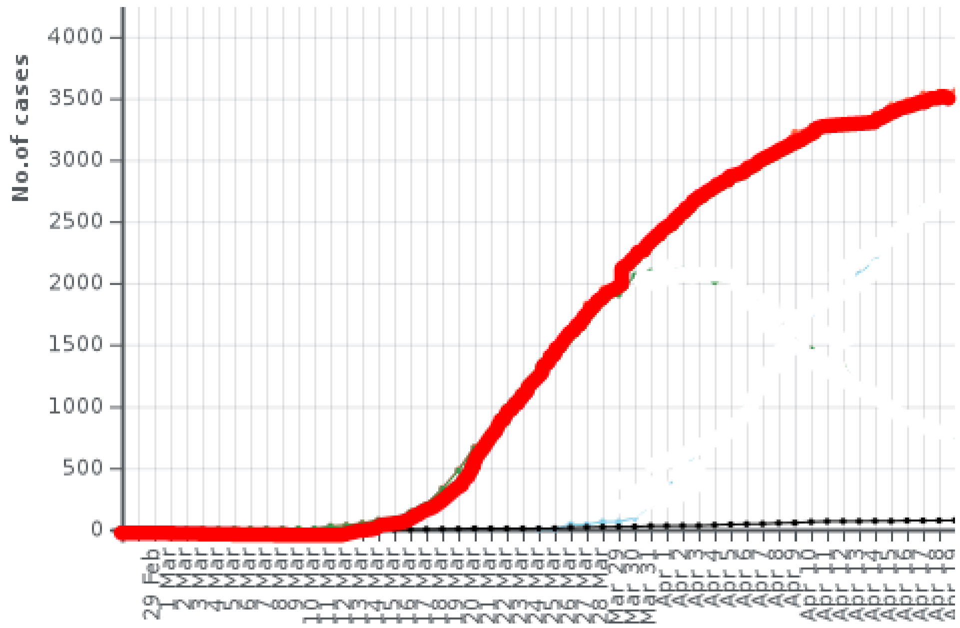
$$A(t) = 1.35567 \cdot 1.46977^t$$

$$B(t) = 12.4396 \cdot t - 34{,}8571$$

Päivänä 16 COVID-19 sairastuneita rekisteröitiin 670 Luxemburgissa.

**i)** **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.

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




**j)** **Anna** kaksi mahdollista syytä siihen, miksi kasvu vähenee voimakkaasti myöhemmässä vaiheessa.

Tätä kuvaajaa voidaan mallintaa funktiolla:

$$C(t) = \frac{3404}{1 + 193 \cdot e^{-0.233 \cdot t}}$$

**k)** **Määritä**, minä päivänä tartuntojen määrän muutos oli suurin.