



Drughunters 2026 – Matematikopgave

Drughunters Medicines Corp. er en (fiktiv) medicinalvirksomhed, der forsøger at udvikle nye typer medicin, der er mere effektive end de eksisterende. For at hjælpe med dette er forskere ved Drughunters Medicines Corp. interesserede i at finde ud af, hvordan og hvorvidt medicin mod migræne fungerer.

Det er rigtig svært at leve med migræne, så det er vigtigt, at der er tilgængelig medicin mod migræne. Dette er selvfølgelig set fra patientens side, men der er også et samfundsperspektiv: op mod hver 7. sygedag i Danmark skyldes migræne, og det koster det danske samfund adskillige milliarder kroner årligt.

Derfor er der en meget stor interesse i at finde den rette behandling mod migræne. Og det er det, Drughunters Medicines Corp. forsker i.

Efter selve opgaveformuleringen er der links til steder, hvor I kan få hjælp til at finde ud af, hvordan I udfører analyserne i f.eks. EXCEL. Der er ikke noget specifikt krav om brug af bestemt software.

Sygdommen og medicinen

- 1) Forklar ganske kort hvad migræne er, og om der findes behandling mod migræne. Hvad er forskellen på akut og præventiv migrænemedicin?

Når man modtager behandling i form af medicin (piller, tabletter, injektioner og lignende), bliver stoffet i kroppen (blodet) en vis mængde tid. Antag, at Drughunters Medicines Corp. har udviklet et præventivt præparat mod migræne, som skal gives ved en injektion en gang om måneden (hver 4. uge). Halveringstiden for dette stof er 18 dage:

- 2) Hvad er halveringstid? Hvor meget stof er tilbage i kroppen lige inden den 2. injektion?
- 3) Bestem en forskrift for den funktion, der angiver indhold af stof i kroppen som funktion af tid. Giv en fortolkning af grundtallet i denne forskrift.



4) Diskutér kort, om det er rimeligt at give en injektion hver 4. uge

Statistisk analyse af data ved hjælp af deskriptiv statistik

Når vi som forskere taler om, at vi skal analysere data, betyder det, at vi ønsker at få et overblik over data og at kunne drage nogle konklusioner vedrørende sammenhænge i data (hvis der overhovedet er nogle). Før vi begynder de mere avancerede matematiske analyser, er det meget vigtigt at bygge noget basisviden op omkring vores data.

I denne opgave får I stillet to ting til rådighed:

- En **Google Colab-notesbog** med kode, der kan generere et syntetisk datasæt med hjælp af en sprogmodel. Her er variablene, der indgår i datasættet forklaret.
- En Excel fil med et syntetisk datasæt for patienter med migræne. I kan antage at datasættet imiterer et virkeligt forsøg med patienter.

I forskning arbejder man ofte med **syntetiske data**, fordi rigtige patientdata er meget følsomme og beskyttet af strenge regler (fx GDPR og etiske retningslinjer). Det betyder, at vi ikke bare kan dele eller bruge rigtige data i undervisning. Med syntetiske data kan vi alligevel bruge de samme metoder, som forskere bruger, uden at afsløre information om virkelige patienter.

- 5) Diskutér hvad fordelene og ulemperne er ved at bruge en sprogmodel til at foreslå variable og sammenhænge. Hvordan kunne man ellers have genereret syntetisk data?
- 6) Ligner de syntetiske data det, man kunne forvente at se hos rigtige migrænepatienter? Hvorfor/hvorfor ikke?
- 7) Overvej (kort!) hvorfor de variable, der er repræsenteret i datasættet, er valgt. Hvad har de med migræne at gøre? Giver de mening? Hvad indeholder variabelen `is_placebo`?



Gå datasættet igennem og lav tabeller med beskrivende/deskriptiv statistik samt grafiske præsentationer af data (gennemsnit, optællinger, figurer osv.) af forskellige variable fra datasættet (evt. mod hinanden) for at få et overblik over data. Hvordan I selekterer (filtrerer/vælger) i data og laver nye variable, kan I finde ud af ved at spørge en AI.

- 8) Lav en ny variabel, der indeholder 'månednummer' for hver række, dvs. 220 rækker med 0'er (nemlig baseline), 220 rækker med 1'ere (month 1) etc. Lav derefter endnu en ny variabel med de tilsvarende målinger af MMD: 220 rækker med baseline_mmd målingerne, 220 rækker med month1_mmd målingerne etc. Således kommer de to nye variable til hver at indeholde 880 rækker. Lav en lineær regressionsanalyse (trendline) med månedsvARIABLEN (0, 1, 2 og 3) på x-aksen og MMD-vARIABLEN på y-aksen. Analysen skal laves således, at den giver to såkaldte regressionslinjer: én for hver behandlingsgruppe (dette kræver, at også variabelen is_placebo er blevet forlænget til 880 rækker).

Angiv og fortolk værdierne af a og b i regressionslinjerne $y = ax + b$ for de to behandlingsgrupper.

- 9) Diskutér om der allerede nu kan etableres nogle konklusioner baseret på de tabeller og grafiske repræsentationer, I har lavet. Skyldes en eventuel sammenhæng korrelation eller kausalitet?

I bestemmer selv, hvor meget beskrivende statistik og hvor mange figurer I vil præsentere for at give et overblik over (sammenhængene mellem) de vigtige variable i jeres datasæt.

Hvordan kan man anvende avancerede dataanalysemetoder til at forstå og forudsige behandlingseffekt ved migræne?

I denne del af opgaven skal I arbejde videre med de data, I har analyseret i statistikdelen, men nu med mere fokus på avancerede metoder inden for Data Science.



Beslutningstræer:

Et beslutningstræ er en model, der bruges til at træffe beslutninger baseret på data. Det er en klassifikationsmetode, der kan opdele data i to eller flere "klasser".

Et beslutningstræ består af noder, der repræsenterer beslutninger eller tests, og grene, der repræsenterer resultaterne af disse beslutninger. Vi arbejder her med et binært beslutningstræ, hvor hver node har to grene: sandt eller falskt, der fører til de næste noder eller til endelige beslutninger (blade). I vores eksempel søger vi at klassificere, om patienter er responderende til en behandling (1) eller ikke-responderende (0) baseret på data fra det forsøg.

I **Google Colab-notesbogen** er der yderligere forklaringer samt kode til at forberede og køre klassifikationen med beslutningstræer.

https://colab.research.google.com/drive/12u0a1fNNgdTb3ILGY_KHcLMMgteKWYl1?usp=sharing

Brug notesbogen til at svare på følgende spørgsmål:

- 10) Forklar med jeres egne ord hvad man bruger beslutningstræer til, og hvordan de konstrueres.
- 11) Hvad kan I tolke ud fra de genererede beslutningstræer?
 - a. Hvad kan være hovedårsagen til, at patienter bliver klassificeret som responderende/ikke responderende
 - b. Er det overraskende? Hvorfor/hvorfor ikke.
- 12) Vi har udeladt nogle variable i vores beregning af beslutningstræerne
 - a. Hvilke variable er udeladt?
 - b. Hvorfor tror I, at netop de variable er udeladt?
- 13) Hvad kunne ellers være interessant at prædiktere enten med det givne data eller med nogle variable, I selv finder på.
 - a. Hvad kan I bruge resultatet til?



I har nu lavet både statistik og data science i denne opgave.

14) Hvad kan I konkludere på baggrund af alle disse spørgsmål, I nu har besvaret?

Generel opgavevejledning

Overordnet set er opgaven opbygget efter følgende model:

- **Spørgsmål 1** omhandler en hjernesygdom og et datasæt. Her handler det primært om at vise, at man er i stand til at udvælge hovedtrækkene og give en så kort og præcis beskrivelse som muligt.
- **Spørgsmål 2-9** omhandler matematiske metoder til grafisk og analytisk at få overblik over og analysere data. Bemærk, at der generelt ikke er noget korrekt facit med to streger under. Alle besvarelser afhænger af de valg, der hele tiden foretages fra jeres side. Det er det, en forsker gør i sit daglige arbejde.
- **Spørgsmål 10-13** omhandler kunstig intelligens til analyser af data og grafisk visualisering og fortolkning af resultaterne.
- **Spørgsmål 14** er en opsummering af, hvad I har lavet.

Vi vil rigtig gerne se jeres posterpræsentation uanset om I har svaret på alle dele af opgaven eller ej.

Til eleverne

Som forsker må man leve med, at der ikke findes endegyldige og korrekte svar. Man må opsøge viden, som andre har skabt eller ved at lave sine egne forsøg. Og så må man med åbent sind holde den viden op imod sin egen videnskabelige hypotese, som derved be- eller afkræftes – eller som oftest kræver yderligere viden for at kunne drage en konklusion. Det kan være en lang og frustrerende proces selv for garvede forskere. Derfor forventer vi selvfølgelig ikke endegyldige løsninger fra jer, men gode forslag hvor der er tænkt over usikkerheder og begrænsninger.

Til finaledagen vil bedømmelseskriterierne være 1/3 formidling og 2/3 faglighed. Det betyder, at det ikke gælder om at have så meget tekst som muligt, men at der skal være et naturligt



flow i fortællingen, så læseren/tilhøreren kan forstå jeres vigtigste pointer. Omvendt er det selvfølgelig heller ikke nok at have en superflot poster, hvis man ikke har svaret på spørgsmålene. Husk at til den mundtlige præsentation behøver I ikke at gennemgå posteren slavisk. Her skal I fokusere på at fremhæve de pointer, som er særligt vigtige for jeres besvarelse. Dommerne har læst posteren på forhånd, men gemmer den endelige bedømmelse til de har set jeres præsentation, hvor de både vil inddrage jeres evne til at fortælle en sammenhængende historie og jeres besvarelse af opfølgende faglige spørgsmål.

Posteren skal ikke nødvendigvis være opdelt således, at I skal gennemgå alle 14 spørgsmål, men kan også være et udvalg af de figurer, tabeller og analyser, I har lavet.

Den skriftlige vurdering er selvfølgelig kun lavet på baggrund af posteren og skal ses som en kort tilbagemelding, ikke en dybtgående analyse af jeres poster.

Rent praktisk skal posteren indsendes som pdf i størrelsen 142x83 cm landskabsformat. Se deadline for indsendelse i kalenderen nedenfor.

Til lærerne

Brug gerne tid i klassen på at snakke om, hvordan hvert enkelt spørgsmål skal forstås, inden I kaster jer over besvarelsen.

Der kan hentes inspiration til, hvordan man kan arbejde med opgaverne på vores hjemmeside Drughunters.

Hints og eksempler på referencer og links (find gerne flere selv)

I skal undervejs benytte EXCEL-funktionen 'Data Analysis'. Hvis I ikke finder den ude til højre under 'Data', kan I gøre følgende:

Tryk File -> Options -> Add-ins -> Analysis ToolPak (VBA) og installér dette



Kalender for Drughunters 2026

2025			2026			
Oktober	November	December	Januar	Februar	Marts	April
	20. okt		31. jan	Tilmelding til Drughunters		
	20. okt	20. dec	Tilmelding til forskerbesøg (max. 20)			
		Forskerbesøg efter aftale		2. feb	31. mar	
	20. okt				27. mar	Opgave- besvarelse
					FINALEDAG	23. apr

Med venlig hilsen
Drughunters 2026