

Eksempel 221:

Stikprøver og databaser

– et forløb indenfor emnet statistik

»Stikprøver og databaser« er et forløb, der introducerer Explorative Data Analysis (EDA) til analyse af spørgeskemaer og stikprøver fra f.eks. databaser. Udgangspunktet er et autentisk datamateriale enten frembragt af eleverne selv, f.eks. via en spørgeskemaundersøgelse, eller hentet på nettet i en passende database. De eksperimentelle data fremstilles grafisk på forskellig vis og der udpeges/undersøges mulige variabelsammenhænge. Forløbet kan eventuelt udvides med en undersøgelse af vigtige statistiske egenskaber for stikprøver i en sådan database.

Mål:

- anvende simple statistiske modeller til beskrivelse af et givet datamateriale, kunne stille spørgsmål ud fra modellen, have blik for, hvilke svar der kan forventes, og være i stand til at formulere konklusioner i et klart sprog.
- anvende it-værktøjer til løsning af givne matematiske problemer.

Niveau:

Projektet egner sig til både til grundforløbet og til højere niveauer.

Samarbejdsmuligheder:

Projektet kan afvikles i matematik, men kan fint indgå i et samarbejde med f.eks. naturvidenskabeligt grundforløb eller samfundsfag C omkring beskrivelse af datasæt. Det vil også afhængigt af det valgte tema kunne indgå i almen studieforberedelse.

Arbejdsformer og tidsforbrug: 8 × 45 minutter.

Klassen introduceres til de begreber og metoder, der er karakteristiske for EDA, dvs. specielt de karakteristiske grafiske værktøjer som f.eks. xy-punktgrafer, kassegrafer og histogrammer til opdagelse af karakteristiske mønstre og sammenhænge, men også til simple kvantitative mål som middelværdier, medianer og kvartiler. Derefter stilles et spørgeskema eller en database til elevernes rådighed for den følgende analyse. Man kan f.eks. udnytte, at man på mange skoler som led i en indledende snak med klassen om deres forudsætninger og forventninger, indleder skoleåret med en spørgeskemaundersøgelse af elevernes fritidsvaner, lektievaneer, forventninger til gymnasiet osv. Hvis disse data opsamles elektronisk, kan de nemt konverteres til en fælles database, hvor oplysningerne om eleverne fremstår i en passende anonymiseret form. Eleverne kan derefter udvælge forskellige størrelser (variable) med henblik på en afklaring af hvilke andre variable, der kan tænkes at påvirke eller blive påvirket af den givne størrelse, ligesom man kan undersøge, om der er forskel mellem forskellige grupper, om der er størrelser, der varierer sammen osv. Undersøgelsen kan munde ud i en skriftlig rapport om de fundne sammenhænge, hvor der sættes ord på tolkningen af graferne, ligesom den kan munde ud i en mundtlig fremlæggelse/diskussion med klassen af de sammenhænge, man mener at have påvist.

Anvendelse af it og værktøjsprogrammer:

Det er oplagt at bruge det statistikværktøj, man alligevel vil benytte i det daglige, f.eks. regnearket Excel eller statistik- og modelleringsprogrammet Fathom.

Undervisningsmaterialer: Man kan finde eksempler på en sådan tilgang til spørgeskemaanalyser i hæftet *Spørgeskemaanalyser*. I dette hæfte refereres til en stor tysk spørgeskemaundersøgelse af gymnasieelevers fritidsvaner Muffins, der er tilgængelig på nettet. På nettet findes desuden adskillige diskussioner af Explorative Data Analysis. F.eks. kan man finde en detaljeret gennemgang EDA fra en naturvidenskabelig synsvinkel i kapitel 1 af håndbogen *Engineering Statistics Handbook*, der frit kan downloades fra nettet stillet til rådighed af det amerikanske National Bureau of Standards.