

Matematikrapport: Statistik

KBJ, december 2023

1s Ma

Opgave 1

Under ”opgave 1” i datafilen ”data.xlsx” findes data for intervalhyppigheden af indlæggelser for to grupper af sygdomme baseret på aldersintervaller. Der er tale om ”lungetyngdomme” og ”øre/næse/hals-sygdomme”. Brug noten ”Deskriptiv statistisk med Nspire” som hjælp til løsning.

- Beregn intervalfrekvens og kumuleret intervalfrekvens for hver af de to dataserier, og bestem gennemsnitsalderen for en indlagt i hver af de to sygdomskategorier.
- Spørgsmålet udgår!*
- Tegn i samme vindue en sumkurve for hver af de to dataserier.
- Bestem kvartilsættene for de to dataserier og bestem ”outlier”-grænserne for hver af dem. Vurdér på den baggrund hvor mange outliers der evt. er i hver dataserie.
- Bestem fra sumkurverne hvilken andel af de indlagte inden for hver af de to sygdomskategorier, som er i den arbejdsduelige alder 18-65 år.
- Skriv en kort tekst der med inddragelse af de statistiske beregninger ovenfor sammenligner aldersprofilen på indlagte inden for hver af de to sygdomskategorier.

Opgave 2

I en undersøgelse har et stort fagforbund undersøgt i en stikprøve blandt 775 af forbundets uaglærte medlemmer hvor mange der lider af allergi. Undersøgelsen viser at det gælder for 204 af de undersøgte. Det vurderes at 21% af den danske befolkning lider af allergi¹.

- Bestem statistisk usikkerhed og 95%-konfidensinterval for den faktiske andel af fagforbundets faglærte medlemmer, som lider af allergi.
- Bestem hvor stor en stikprøve fagforbundet skulle have brugt, for at det med den fundne stikprøveandel havde givet en statistisk usikkerhed på højst 2%-point.
- Hvad kan fagforbundet konkludere når undersøgelsen holdes op imod andelen af alle danskere?

¹ <https://www.benadryl.dk/allergi/hvor-mange-har-allergi>

Opgave 3

I ”Opgave 3” i filen ”data.xlsx” findes målinger af højde og hvilepuls målt for 320 unge mænd i forbindelse med vurdering af egnethed til militær tjeneste (session).

- Bestem middelværdi og spredning for såvel højde som hvilepuls for de 320 unge mænd.
- Bestem det udvidede kvartilsæt for hver af de to dataserier og undersøg hvor mange outliers de to serier indeholder, samt om de evt. er højre- eller venstreskæve. Inddrag tegning af boksplot.
- Undersøg ved lineær regression om der er en statistisk sammenhæng mellem ”højde” og ”hvilepuls”. Inddrag et residualplot i vurderingen.
- Opstil ud fra den lineære regression den bedste model $f(x) = a \cdot x + b$ hvor x angiver højde (målt i cm) og $f(x)$ angiver hvilepuls (målt i slag pr. sekund).
- Bestem største positive og negative absolutte afvigelse (residual), og bestem hvilken relativ afvigelse de to afvigelser hver især svarer til.
- Benyt modellen til at bestemme den forventede hvilepuls af en ung mand med en højde på hhv. 200 og 140 cm.
- Benyt modellen til at bestemme den forventede højde af en unge mand med en hvilepuls på 65.

GRUPPER TIL ARBEJDET (aflever én besvarelse pr. gruppe)

Gruppe 1 Kerem Simon Frederik Nikolaj	Gruppe 2 Margrethe Mathilde C.N. Lily Trishna	Gruppe 3 Daniel Mikkel Lauritz Isak	Gruppe 4 Mathilde D.N. Melika Nikita Victoria
Gruppe 5 Bedros Mads Mathias Ibrahim	Gruppe 6 Isabella Sarah Kristine Ella	Gruppe 7 Thor Jeppe Aske Bogdan	