



Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4245 Statistikk
Vår 2015

Øving nummer 7, blokk II (Matlab-øving 2)

Oppgave 1

I denne oppgaven skal vi analysere tre datasett. Dataene er daglige temperaturobservasjoner for Trondheim, Tynset og Bodø i 2013. Vi har dermed 365 observasjoner i hvert datasett. Merk at mye i denne oppgaven baserer seg på det som diskuteres i kapittel 1 i læreboka.

Vi skal starte med å laste dataene inn i Matlab.

- a) Datasettene kan du hente ut fra filene `trondheim.txt`, `bodo.txt` og `tynset.txt` som er tilgjengelige på hjemmesiden.

Vi laster dataene inn i MATLAB ved å bruke funksjonen `load`.

```
trondheim = load('trondheim.txt');  
bodo = load('bodo.txt');  
tynset = load('tynset.txt');
```

Temperaturobservasjonene finner vi i kolonne 6 i matrisene `trondheim`, `baatsfjord` og `tynset`. Vi kan opprette egne vektorer med temperaturobservasjonene,

```
TRD = trondheim(:,6);  
BO = bodo(:,6);  
TY = tynset(:,6);
```

Vi har nå tre vektorer i MATLAB; `TRD`, `BO` og `TY`. Det kan være kjekt å samle alle dataene i en og samme matrise. Det gjør du ved å skrive

```
MT = [TRD BO TY];
```

`MT` er nå en matrise der første, andre og tredje kolonne er gjennomsnittstemperaturer for hver dag i året i henholdsvis Trondheim, Bodø og på Tynset.

Vi anbefaler deg i det følgende å bruke MATLAB til å utføre analyse av datasettene. Bruk `help` i MATLAB for å lære om funksjonene du skal bruke. For eksempel kan du skrive `help mean` for å finne ut av hva funksjonen `mean` gjør i MATLAB.

MATLAB tips: Noen relevante MATLAB funksjoner er `mean`, `median`, `std` og `var`.

- b) Utfør en beskrivende analyse av de tre datasettene og lag to relevante plott/histogram for hvert datasett. Sammenlign forholdet mellom gjennomsnittsverdi og empirisk median i hvert datasett med forholdet mellom gjennomsnittsverdi og median for utfall av en normalfordeling. Er resultatet slik du ville forvente? For hint til relevante plott og MATLAB-funksjoner, se øving 1.

- c) Presenter alle datasettene med boksplott. Hvilke konklusjoner kan du trekke basert på disse boksplottene?

MATLAB tips: Funksjonen du trenger er `boxplot`. Bruker du matrisen MT som input i denne funksjonen får du alle boksplottene i samme figur.

- d) Er det noen ekstremverdier i datasettene? Hvilken metode bruker MATLAB for å bestemme hvorvidt en observasjon er en ekstremverdi eller ikke?
- e) Er det noen forskjeller i de tre boksplottene som tyder på at datasettene kommer fra populasjoner med forskjellige fordelinger?
- f) Lag et normal kvantil-kvantil plott (QQ-plott) for å evaluere om datasettene kommer fra en normalfordeling eller ikke.

MATLAB tips: Funksjonen du trenger er `normplot`.

- g) Lag minst tre scatterplott der du plotter temperaturobservasjonene fra de tre stedene mot hverandre.

MATLAB tips: Funksjonen du trenger er `scatter`.

- h) Bruk MATLAB til å regne ut korrelasjonsmatrisen for de tre datasettene.

MATLAB tips: Funksjonen du trenger er `corr`.

- i) Er datasettene positivt korrelerte? Er dette noe du ville forventet? Forklar.

Oppgave 2

I denne oppgaven skal vi simulere utfall fra en kjent sannsynlighetsfordeling. Disse utfallene skal vi bruke til å undersøke effekten av sentralgrenseteoremet på fordelingen av gjennomsnittsverdien til utfallene.

I fila SGT.m som er tilgjengelig fra hjemmesiden finner du kode for å utføre lignende oppgaver som du skal gjøre her.

- a) Simuler 1000 datasett i MATLAB. Hvert datasett skal bestå av 100 utfall fra en normalfordeling med forventningsverdi 5 og standardavvik 2.

MATLAB tips: Bruk funksjonen `normrnd`.

- b) Regn ut gjennomsnittsverdien av alle de 1000 datasettene. Lag et histogram basert på gjennomsnittsverdiene du har regnet ut. Minner formen på histogrammet om formen til en normalfordeling? Var dette forventet? Forklar.

MATLAB tips: Bruk funksjonene `hist` og `normplot`.

- c) Gjør det samme som i a), men nå skal utfallene komme fra en binomisk fordeling, $\text{Bin}(N,p)$, $N = 5, p = 0.2$. Prøv deg frem med $n = 2, 5, 10, 20, 30, 50, 100$ utfall for hvert datasett.

MATLAB tips: Bruk funksjonen `binornd`.

- d) Hvilke av simuleringene gir et histogram som ligner en normalfordeling? Bruk sentralgrenseteoremet til å forklare resultatet du får.

MATLAB tips: Bruk funksjonene `hist` og `normplot`.