

Statistikk for bergvarmeanlegg

Oppvarming og kjøling av bygninger er en av de tre store driverne av klimagassutslipp i verden, sammen med transport og industri. I det grønne skiftet trengs det grønne løsninger for bygninger, og gjerne løsninger som belaster elektrisitetsnettet minst mulig. Varmepumper vil være en del av dette bildet, men som vi nordboere vet går virkningsgraden til varmpumper ned når det blir kaldt ute og vi trenger varmen som mest. Bergvarme er en mulig løsning på dette problemet. Dypt nede holder grunnfjellet en jevn temperatur på rundt 4 °C året rundt. Energibrønner boret ned i grunnfjellet gir altså varmepumpen en jevn temperatur å arbeide mot hele året.

Dataene som brukes i denne oppgaven kommer fra hotellet Scandic Flesland Airport på Flesland i Bergen. Vi skal se på et datasett for 2018 som inneholder data for mengden varme som er hentet ut til romoppvarming og mengden varme som ble ført tilbake i berget ved å kjøle ned inneluften om sommeren.

Datafilene viser kumulativt energiforbruk siden oppstarten av anlegget, logget med en periode på 10 min.

1. Installer Python-biblioteket pandas og bruk det til å laste inn datafilene `varme.csv` og `kulde.csv` ved hjelp av funksjonen `pandas.read_csv`.

Hint: Den første kolonnen i hver av datafilene inneholder tidspunkter. Det kan være nyttig å bruke argumentet `parse_dates = ['Timestamp']` for at pandas skal tolke denne kolonnen som tidspunkter i stedet for strenger.

2. Hva er totalforbruket av romvarme for 2018? Hva er totalforbruket for hver av årets 12 måneder? Vis resultatene som en graf.

Hint: pandas bruker en egen datatype som kalles *datarammer*. Bruk en for-løkke til å løpe over alle linjene i datarammen når du skal finne totalforbruket for hver av årets måneder. Du kan adressere en enkeltlinje `i` i datarammen `df` ved hjelp av syntaksen `df.iloc[i]`. Verdien i kolonne `j` kan hentes ut med `df.iloc[i, j]`. Månednummeret til et tidspunkt `time` er lagret i egenskapen `month`; syntaksen `time.month` gir altså månednummeret. Når du tegner grafen kan det være nyttig å bruke funksjonen `matplotlib.pyplot.plot`.

Vi skal nå se litt nærmere på månedene januar og juli.

3. Hva er forbruket av romvarme for hver av dagene i disse månedene? Inspiser dagsforbruket for hver av disse to månedene ved hjelp av et spredningsplott. Finnes det noen ekstreme verdier (*outliers*) i disse to månedene?

Hint: Syntaksen `time.day` gir deg dagnummeret til et tidspunkt. Funksjonen `matplotlib.pyplot.scatter` kan brukes til å lage spredningsplott.

4. Hva er gjennomsnittlig dagsforbruk av romvarme disse to månedene?
Hva er standardavviket?

Hint: Gjennomsnitt og standardavvik kan regnes ut ved hjelp av funksjonene `numpy.mean` og `numpy.std`.

5. Funksjonen `numpy.std` tar argumentet `ddof`. Hvorfor må du gi argumentet `ddof = 1` når du skal bruke funksjonen til å regne ut standardavviket?

Hint: Slå opp i dokumentasjonen til funksjonen på <https://numpy.org/doc/stable/> hvis du lurer på hva argumentet `ddof` gjør.

6. Presenter forbruket i månedene januar og juli som et boksplott.

Hint: Bruk funksjonen `matplotlib.pyplot.boxplot` til å tegne et boksplott.

Nå skal vi se på det daglige forbruket av romvarme i hele 2018.

7. Lag et histogram som viser forbruket av romvarme.

Hint: Bruk funksjonen `matplotlib.pyplot.hist` til å tegne et histogram.

En vanlig grafisk beskrivelse av energiforbruket til et bygg er *varighetskurven*. Varighetskurven har energiforbruk pr. dag på y -aksen og årets dager (1–365) på x -aksen. Den viser antallet dager der energiforbruket er større eller lik en gitt verdi y .

8. Lag varighetskurver for romoppvarming og kjøling.

Hint: Bruk den innebygde funksjonen `sorted` til å sortere listene med dagsforbruket av romvarme og kjøling. Med argumentet `reverse = True` får du en liste som er sortert i synkende rekkefølge.

9. Hvilke likheter og forskjeller ser du mellom varighetskurver og persentilplott?
10. Hva forteller disse varighetskurvene? Hvilke sammenheng ser du mellom varighetskurven for romoppvarming og histogrammet fra oppgave 7?

Kvantilplott er et nyttig hjelpemiddel for å avgjøre grafisk hva slags fordeling et datasett har.

11. Lag et kvantilplott for romoppvarming ved hjelp av en standard nor-

malfordeling ($\mu = 0, \sigma^2 = 1$). Er forbruket av romvarme normalfordelt? Sammenlign med histogrammet fra oppgave 7.

Hint: Funksjonen

$$q_{0,1}(f_i) = 4.91[f_i^{0.14} - (1 - f_i)^{0.14}]$$

er en god tilnærming til kvantilene til en standard normalfordeling. Her er $f_i = (i - \frac{3}{8})/(n + \frac{1}{4})$, der i er nummeret til elementet i utvalget og n er det totale antallet elementer i utvalget.

Vi begrenser oss nå til å se på romoppvarming i månedene mai–september, og skal anta at dagsforbruket i er normalfordelt i disse månedene.

12. Tegn et nytt histogram og et nytt kvantilplott for dette utvalget for å teste denne antakelsen.
13. Hva er gjennomsnittlig dagsforbruk? Finn et 95% konfidensintervall for forventningsverdien til dagsforbruket. Variansen til normalfordelingen er ikke kjent.

Hint: Biblioteket `scipy.stats` inneholder mange nyttige statistikk-verktøy. Objektene `scipy.stats.norm`, `scipy.stats.t` og `scipy.stats.chi2` gir deg henholdsvis en normalfordelt, t -fordelt og χ^2 -fordelt stokastisk variabel. Alle disse objektene inneholder flere nyttige funksjoner. For eksempel gir funksjonen `pdf` gir deg tetthetsfunksjonen til fordelingen, funksjonen `cdf` den kumulative tetthetsfunksjonen og `ppf` gir deg persentiler av fordelingen. Bruk `scipy.stats.t.ppf` til å finne konfidensintervallet for gjennomsnittet.

14. Finn et 95% konfidensintervall for variansen til dagsforbruket.

Hint: Bruk funksjonen `scipy.stats.chi2.ppf`.

Det er interessant å se på disse dataene i lys av værdata for Flesland.

15. Last inn data for døgnmiddeltemperaturen på Flesland ved hjelp av `pandas` og tegn grafene til forbruket av romvarme og -kjøling for hele 2018 som funksjon av utendørstemperaturen.

Energiberegningene som ble gjort i forkant av byggingen forutsa at hotellet ville ha behov for kjøling når utendørstemperaturen steg over 15 °C.

16. Bedøm visuelt om dette stemmer basert på grafen fra forrige oppgave.

Varighetskurven til bygget viser at energibehovet til bygget er dominert av oppvarming. Jo bedre isolert bygget er, jo lavere vil altså energiregningen bli. Samtidig vil god isolering bety at bygget behøver kjøling ved en lavere

temperatur. Nå skal vi teste om bygget faktisk er så godt til å holde på varmen som det var planlagt å være ved å se på kjøledataene.

Nullhypotesen vår er at bygget trenger kjøling 95% av alle dager der $T > 15$ °C. Som det framgår av grafen fra oppgave 13 er det noe kjølebehov selv helt ned mot 0 °C for å utligne temperaturforskjeller i bygget. Vi velger oss derfor en terskelverdi på 0.5 MWh/dag for å definere når det er et reelt behov for kjøling. Nullhypotesen vår er altså at $p = 0.95$ for at kjølebehovet er større enn 0.5 MWh når $T > 15$ °C. Alternativhypotesen er at bygget er dårligere isolert, det vil si at $p < 0.95$. For å unngå å forkaste nullhypotesen selv om den er riktig lar vi området der vi forkaster nullhypotesen begynne ved $p = 0.90$.

17. Hvor mange punkter i datasettet har $T > 15$ °C? Hvor mange punkter med kjølebehov mindre enn 0.5 MWh/dag trenger vi for å forkaste nullhypotesen? Støtter datasettet nullhypotesen?

18. Hva er signifikansnivået til denne testen?

Hint: Biblioteket `scipy.stats` inneholder også objektet `scipy.stats.binom`. Funksjonen `scipy.stats.binom.pmf` gir deg sannsynlighetsmassen til den diskrete fordelingen.

19. Tegn en graf som viser sannsynligheten for å gjøre en type II-feil som funksjon av verdien p for $0.75 < p < 0.90$.