

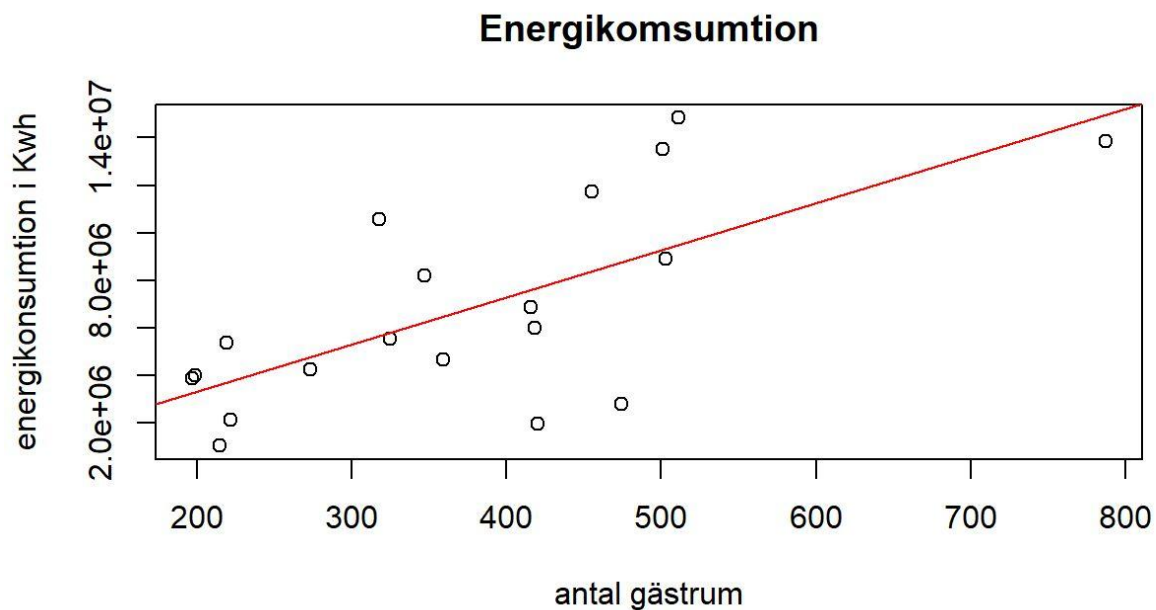
Enkel linjär regression

- Hotel energi

Linda Li Bogardi

I denna rapport ska vi undersöka hur energikonsumtion (Energy Consumption in Kilowatt-hours) förhåller sig till antalet gästrum (number of guest rooms). Vi väljer antalet gästrum som förklarande variabel (x) och energikonsumtion som responsvariabel (y). Vi ska genom att beräkna korrelationen mellan ovan nämnda parametrar, ta ut en linjär regressionsmodell, anpassa en linjär formel och besvara vår frågeställning.

Vår hypotes är: det finns ett positivt linjärt samband mellan energikonsumtion och antalet gästrum. Ju mer gästrum hotellet har, dessutom högre blir energikonsumtion.



Figur 1. Energikonsumtion plots mot antalet gästrum.

$$y_i = m + kx_i + \varepsilon_i$$

Vi antar att y_i (energikonsumtion) är i samband med x_i (antal gästrum) linjärt som beskrivs i ovanstående formel. Där ε_i är oberoende slumpvariabler och $\varepsilon_i \sim N(0, \sigma^2)$.

Genom anpassning av det stickprovsdata som plottas i Figur 1, får vi skattade parametrar:

$$m = -612721.94 \approx -0.6e + 6$$

$$k = 19802.45 \approx 0.02e + 6$$

Det innebär att när $x = 0$, dvs. när det inte finns någon hotellrum, vinner man 0,6 Mega Kwh energi! Det stämmer inte. m borde vara minst noll.

k -värdet är 19802.45, men om vi skriver om värdet med prefix Mega så får vi 0,02. Det betyder att vi skulle kunna plotta $y_i \sim x_i$ då y_i med prefix Mega så får vi $k \approx 0,02$. k -värdet är inte noll, vilket tyder på att x och y är linjärt korrelerade. Vi kan besvara vår frågeställning nu att energikonsumtionen förhåller sig till antal gästrum linjärt med positiv lutning. Dvs. att ökat antal gästrum har ett samband till ökad energikonsumtion. Dock innebär det inte nödvändigtvis att större antal gästrum är den direkta orsaken till höjd energikonsumtion. Vår resultat stämmer väl med vår hypotes.

Kovariansen är $0,685091 \approx 0,69 > 0$, vilket tyder på att våra två parametrar är korrelerade, och när antalet gästrum ökar, så ökar energikonsumtionen.

F-statistic: 15.04 on 1 and 17 DF, p-value: 0.001209

Vi har ett högt F-värde och ett lågt p-värde, och det innebär att linjär regressionsmodell är statistiskt signifikant.