Övning 3 – Statistisk översiktskurs (del 1 – relaterar till föreläsning 7)

Korrelation

n – antalet observationer

i – indexerar observationerna

Antag att du har två variabler, x och y

Standardavvikelse(x) = $s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Standardavvikelse(y) = $s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

r = Korrelationskoefficient för sambandet mellan x och y = $\frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y}$ (från F7)

Uppgift 3.1

Normalt tar vi inte fram korrelationskoefficienten manuellt, men det kan vara bra att gå igenom någon sådan beräkning.

Antag att du har följande data, du observerar x och y för tre (n=3) olika objekt (dina observationer) och får följande värden.

$\{x_1, x_2, x_3\} = \{2, 3, 4\}$

$\{y_1, y_2, y_3\} = \{2, 2, 5\}$

Båda sekvenserna har medelvärde 3, dvs $\bar{x} = 3, \bar{y} = 3$.

A) Rita ett spridningsdiagram med de tre datapunkterna

B) I övning 1 tog du fram standardavvikelsen för sekvensen {2, 3, 4}, vi fick s_x=1. Ta fram standardavvikelsen för sekvensen {2, 2, 5}, du ska få svaret s_y=$\sqrt{3}$

C) Ta fram korrelationskoefficienten:

r = $\frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y}$ = $\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y}$ = $\frac{1}{2s_xs_y}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$=
$\frac{1}{2\times1\times\sqrt{3}}((2-3)(2-3) + (3-3)(2-3) + (4-3)(5-3))$=
$\frac{1}{2\times\sqrt{3}}((-1)(-1) + (0)(-1) + (1)(2))$=
$\frac{1}{2\times\sqrt{3}}(1 + 0 + 2)$=$\frac{1}{2\times\sqrt{3}}(3) = \frac{3}{2\times\sqrt{3}} \approx 0.866$

Vi får en hög korrelation, nära 1, gå tillbaka till ditt spridningsdiagram i A och gå igenom om du tycker värdet känsligt rimligt.