Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Kovarians og korrelasjon

$\mu_X = E(X)$, $\mu_Y = E(Y)$. $\sigma_X^2 = Var(X)$, $\sigma_Y^2 = Var(Y)$.
Kovariansen til stokastiske variable X og Y er

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X \mu_Y,$$

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

Dersom stokastiske variable X og Y er uavhengige er korrelasjonen lik 0.
Da er $E(XY) = E(X)E(Y) = \mu_X \mu_Y$.

Regneregel for sum

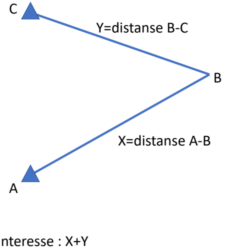
$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

Dersom variablene er uavhengige så forsvinner siste ledd.
Kovariansen er 0.

Eksempel: Oppmåling av to distanser

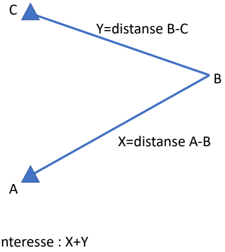
Jonas måler opp distansen X fra A til B, og Y fra B til C. Han er interessert i distansen fra A til C via B: $X + Y$.



$$E(X + Y) = E(X) + E(Y)$$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Eksempel: Oppmåling av to distanser



$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

- Med varians $\sigma_X^2 = \sigma_Y^2 = 1$ og uavhengige målinger, blir variansen lik 2.
- Med varians $\sigma_X^2 = \sigma_Y^2 = 1$ og korrelasjon 0.9 mellom målinger, blir variansen lik 3.8.
- Med varians $\sigma_X^2 = \sigma_Y^2 = 1$ og korrelasjon -0.9 mellom målinger, blir variansen lik 0.2.

Forventning og varians til linearkombinasjoner

Anta stokastiske variable $X_1, \ldots, X_n$.
Sett $\mu_i = E(X_i)$, $\sigma_i^2 = Var(X_i)$.
$Y = \sum_{i=1}^n a_i X_i$
Da er

$$E(Y) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) = \sum_{i=1}^n a_i \mu_i$$

Dersom variablene er **uavhengige** har vi:

$$Var(Y) = Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var(X_i) = \sum_{i=1}^n a_i^2 \sigma_i^2,$$

(Dersom de ikke er uavhengige, må vi ta med alle kovariansene mellom X_i og X_j.)

Spesialtilfelle

Anta stokastiske variable $X_1, \ldots, X_n$.
Sett $\mu_i = E(X_i) = \mu$, $\sigma_i^2 = Var(X_i) = \sigma^2$.
Gjennomsnitt: Sett $a_i = 1/n$.
$Y = \sum_{i=1}^n a_i X_i = \frac{1}{n} \sum_{i=1}^n X_i$
Da er

$$E(Y) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

Dersom variablene er uavhengige har vi:

$$Var(Y) = \sum_{i=1}^{n} a_i^2 Var(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n},$$

Spesialtilfelle plot

Gjennomsnittet $E(Y) = 0$ og

$$Stdev(Y) = \sqrt{Var(Y)} = \frac{\sigma}{\sqrt{n}},$$