Tõenäosusteooria ja statistika II

Loeng 4

Dispersioonist ja kovariatsioonist

Märt Möls
mart.mols@ut.ee

Dispersioon, $\mathrm{D}(X) = \mathrm{E}\left(X - \mathrm{E}(X)\right)^2$

$$= \mathrm{E}\left[(X - \mathrm{E}(X))\,(X - \mathrm{E}(X))\right]$$

Omadused (T&MS I):

1. $\mathrm{D}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$
2. $\mathrm{D}(c) = 0$
3. $\mathrm{D}(c + X) = \mathrm{D}(X)$
4. $\mathrm{D}(cX) = c^2\mathrm{D}(X)$
5. $\mathrm{D}(X + Y) = \mathrm{D}(X) + \mathrm{D}(Y) + 2\mathrm{cov}(X,Y)$
6. Kui X ja Y on sõltumatud, siis

$$\mathrm{D}(X + Y) = \mathrm{D}(X) + \mathrm{D}(Y)$$

$\mathrm{cov}(X,Y) = \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right]$

Näide 1: valimikeskmise dispersioon

$$\mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \overset{(4)}{=} \frac{1}{n^2}\mathrm{D}\left(\sum_{i=1}^{n} X_i\right) \overset{(6)}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\left(X_i\right) = \frac{1}{n^2}n\mathrm{D}\left(X\right) = \frac{\mathrm{D}\left(X\right)}{n}$$

Näide 2: valimikeskmise dispersioon

$$\mathrm{cov}(X_{2i-1}, X_{2i}) = \mathrm{D}(X), \qquad i = 1\ldots n/2$$

$$X_{2i} \perp X_j, \qquad j \neq 2i-1, j \neq 2i$$

$$\mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \ldots$$

Juhuslike suuruste X ja Y kovariatsioon on defineeritud kui

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathrm{E}\left[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\right]\\ &= \mathrm{E}\left[XY - X\mathrm{E}(Y) - Y\mathrm{E}(X) + \mathrm{E}(X)\mathrm{E}(Y)\right]\\ &= \mathrm{E}\left(XY\right) - \mathrm{E}\left(X\mathrm{E}(Y)\right) - \mathrm{E}\left(Y\mathrm{E}(X)\right) + \mathrm{E}(X)\mathrm{E}(Y)\\ &= \mathrm{E}\left(XY\right) - \mathrm{E}(Y)\mathrm{E}\left(X\right) - \mathrm{E}(X)\mathrm{E}\left(Y\right) + \mathrm{E}(X)\mathrm{E}(Y)\\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

Juhuslike suuruste X ja Y kovariatsioon on defineeritud kui

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathrm{E}\left[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\right]\\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

$\operatorname{cov}(X,X) = \mathrm{D}(X)$

$\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$

Kui $X \perp Y$, siis $\operatorname{cov}(X,Y) = 0$

$\operatorname{cov}(c + X,Y) = \operatorname{cov}(X,Y)$

$\operatorname{cov}(a\,X + b\,Y, Z) = a\operatorname{cov}(X,Z) \;+\; b\operatorname{cov}(Y,Z)$

Juhusliku vektori keskväärtus ja dispersioon

Juhusliku vektori

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

keskväärtus

$$\mathrm{E}(\mathbf{X}) = \begin{pmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_n) \end{pmatrix}$$

Analoogselt on defineeritud ka juhusliku maatriksi keskväärtus

dispersioon

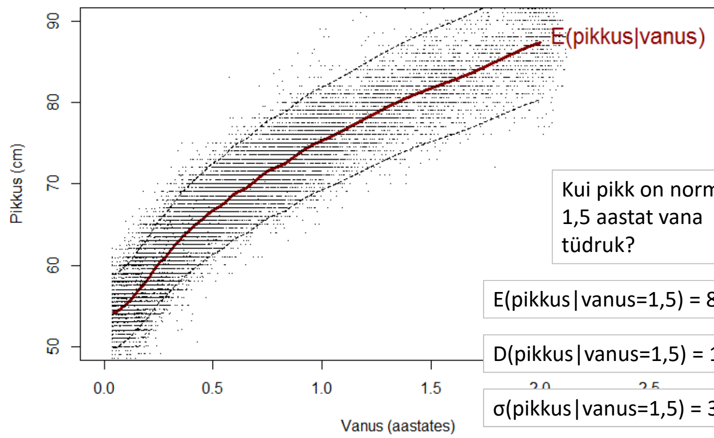
$$\mathrm{D}(\mathbf{X}) = \mathrm{E}\left[(\mathbf{X}-\mathrm{E}(\mathbf{X}))(\mathbf{X}-E(\mathbf{X}))^T\right]$$

$$\mathrm{D}(\mathbf{X}) = \begin{pmatrix} \mathrm{D}X_1 & \operatorname{cov}(X_1,\ X_2) & \mathrm{cov} \\ \operatorname{cov}(X_2,\ X_1) & \mathrm{D}X_2 & \mathrm{cov} \\ \operatorname{cov}(X_3,\ X_1) & \operatorname{cov}(X_3,\ X_2) & \\ \vdots & \vdots & \\ \operatorname{cov}(X_n,\ X_1) & \operatorname{cov}(X_n,\ X_2) & \mathrm{cov} \end{pmatrix}$$

Tinglik dispersioon, D(*Y*|*X*=*x*)

Juhusliku suuruse *Y*|*X*=*x* dispersioon

$$\begin{aligned}\mathrm{D}(Y|X=x) &= \mathrm{E}\left[\left(Y-\mathrm{E}(Y|X=x)\right)^2|X=x\right]\\ &= \mathrm{E}\left(Y^2|X=x\right) - \left[\mathrm{E}(Y|X=x)\right]^2\end{aligned}$$



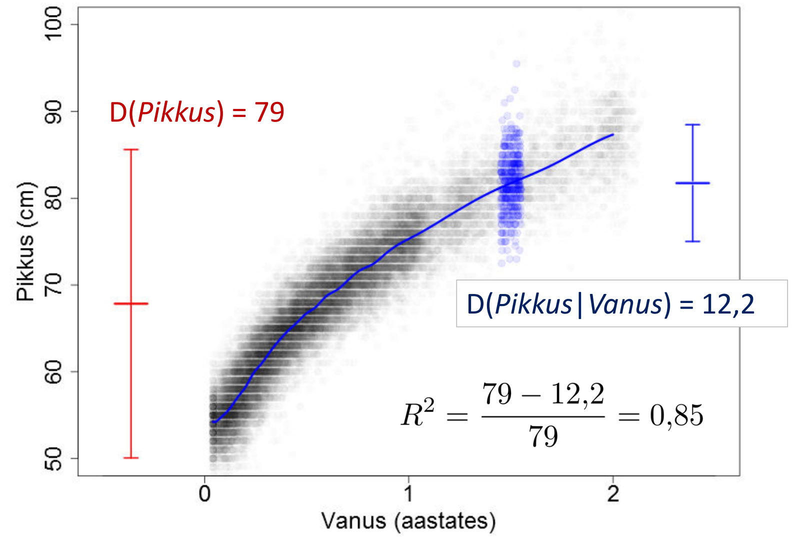
Kui pikk on normaalne 1,5 aastat vana tüdruk?

E(pikkus|vanus=1,5) = 81,7 cm

D(pikkus|vanus=1,5) = 12,2 cm

σ(pikkus|vanus=1,5) = 3,5 cm

Ootuspärane: 74,7 … 88,7



$$\mathrm{D}\left(\mathrm{E}(Y|X)\right) = \mathrm{E}\left([E(Y|X)]^2\right) - \underbrace{\left(\mathrm{E}\mathrm{E}(Y|X)\right)^2}_{(\mathrm{E}Y)^2}$$

$$\mathrm{E}\left(\mathrm{D}(Y|X)\right) = \underbrace{\mathrm{E}\left(E(Y^2|X)\right)}_{\mathrm{E}\left(Y^2\right)} - \mathrm{E}\left([\mathrm{E}(Y|X)]^2\right)$$

$$\mathrm{D}(Y) = \mathrm{D}\left(\mathrm{E}(Y|X)\right) + \mathrm{E}\left(\mathrm{D}(Y|X)\right)$$

Näide 3: valimikeskmise dispersioon

$$\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \dots$$

$$\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i \middle| N\right) = \mathrm{E}(X) \qquad \mathrm{D}\left(\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i \middle| N\right)\right) = 0$$

$$\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N} X_i \middle| N\right) = \frac{1}{N^2}\mathrm{D}\left(\sum_{i=1}^{N} X_i \middle| N\right) = \frac{1}{N^2} N\mathrm{D}\left(X\right)$$

$$= \frac{\mathrm{D}\left(X\right)}{N}$$

$$\mathrm{E}\left[\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N} X_i \middle| N\right)\right] = \mathrm{E}\left(\frac{\mathrm{D}\left(X\right)}{N}\right) = \mathrm{E}\left(\frac{1}{N}\right)\mathrm{D}\left(X\right)$$

Näide 3: valimikeskmise dispersioon

$$\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)=\mathrm{E}\left(\frac{1}{N}\right)\mathrm{D}\left(X\right)$$

$$\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}X_i\middle|N\right)=\mathrm{E}(X) \qquad \mathrm{D}\left(\mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}X_i\middle|N\right)\right)=0$$

$$\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N}X_i\middle|N\right)=\frac{1}{N^2}\mathrm{D}\left(\sum_{i=1}^{N}X_i\middle|N\right)=\frac{1}{N^2}N\mathrm{D}\left(X\right)$$

$$=\frac{\mathrm{D}\left(X\right)}{N}$$

$$\mathrm{E}\left[\mathrm{D}\left(\frac{1}{N}\sum_{i=1}^{N}X_i\middle|N\right)\right]=\mathrm{E}\left(\frac{\mathrm{D}\left(X\right)}{N}\right)=\mathrm{E}\left(\frac{1}{N}\right)\mathrm{D}\left(X\right)$$

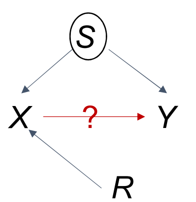
Lineaarne maailm…

$$Y=\beta_1+\beta_{XY}X+\varepsilon_Y$$

$$\begin{aligned}&\mathrm{E}X\\ &\mathrm{E}Y=\beta_1+\beta_{XY}\mathrm{E}X\\ &\beta_1=\mathrm{E}Y-\beta_{XY}\mathrm{E}X\end{aligned}$$

$$\begin{aligned}\mathrm{cov}(X,Y)&=\mathrm{cov}(X,\beta_1+\beta_{XY}X+\varepsilon_Y)\\&=\mathrm{cov}(X,\beta_{XY}X)\\&=\beta_{XY}\mathrm{D}(X)\\ \beta_{XY}&=\mathrm{cov}(X,Y)/\mathrm{D}(X)\end{aligned}$$

$$\begin{aligned}\mathrm{D}(Y)&=\beta_{XY}^2\mathrm{D}(X)+\mathrm{D}(\varepsilon_Y)\\ \mathrm{D}(\varepsilon_Y)&=\mathrm{D}(Y)-\beta_{XY}^2\mathrm{D}(X)\end{aligned}$$



Lineaarne maailm…

$$\begin{aligned}Y&=\beta_1+\beta_{SY}S+\beta_{XY}X+\varepsilon_Y\\ X&=\beta_2+\beta_{SX}S+\beta_{RX}R+\varepsilon_X\end{aligned}$$

Tunnus *R* on **instrument**, kui:

a) ta mõjutab tunnuse *X* väärtuste kujunemist;
b) ta võib mõjutada tunnust *Y* vaid tunnuse *X* kaudu – teisi otseseid või kaudseid võimalusi (teid) mõjutada tunnuse Y väärtuseid tal pole – ja puuduvad ka tunnused, mis mõjutaksid nii R-i kui ka Y-it (või R'i ja X'i) samaaegselt (segavad faktorid)

$$\begin{aligned}\mathrm{cov}(R;Y)&=\mathrm{cov}(R;\beta_{XY}\beta_{RX}R)\\&=\beta_{XY}\beta_{RX}\mathrm{D}R\end{aligned}$$

$$\begin{aligned}\mathrm{cov}(R;X)&=\mathrm{cov}(R;\beta_{RX}R)\\&=\beta_{RX}\mathrm{D}R\end{aligned}$$

$$\beta_{XY}=\mathrm{cov}(R;Y)/\mathrm{cov}(R;X)$$

$$\hat{\beta}_{XY}=\widehat{\mathrm{cov}}(R;Y)/\widehat{\mathrm{cov}}(R;X)$$