

Normaaljaotus

Valim ja populatsioon standardviga

5. loeng

Normaaljaotus

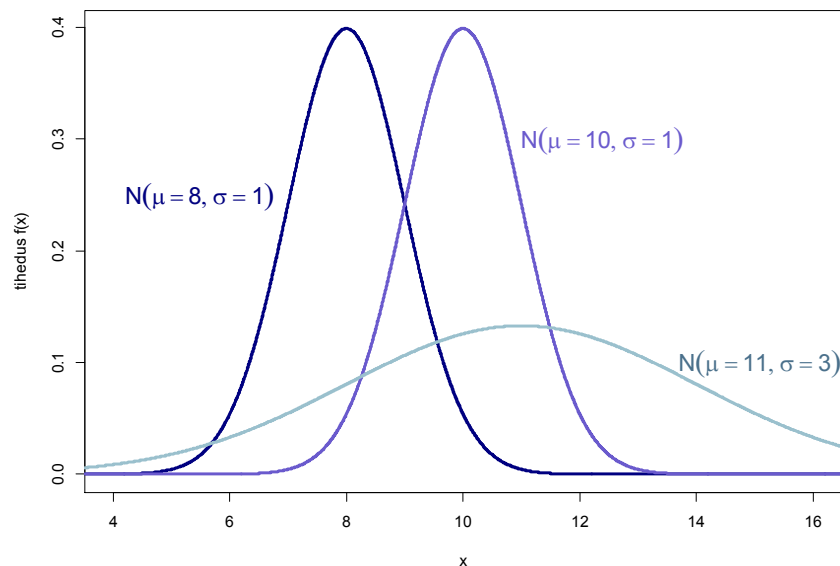
Üks sagedamini ette tulev jaotus eluslooduses. Kui uuritavat tunnust mõjutavad paljud erinevad tegurid, millest ühegi mõju pole omaette võttes märkimisväärne, siis on uuritava tunnuse jaotus lähedane normaaljaotusele. Seega on näiteks normaaljaotusega väga paljude geenide poolt määratavad näitajad: pikkus, lehmade piimaand, lihakaal, 100 tera mass,

Kui uuritava tunnuse jaotus on normaaljaotusega, siis tema tihedusfunktsioon on esitatav järgmisel kujul:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

Kus μ on uuritava tunnuse keskvärtus ($\sim$ uuritava tunnuse kõigi väärtuste keskmine) ja σ^2 on uuritava tunnuse dispersioon.

Normaaljaotus



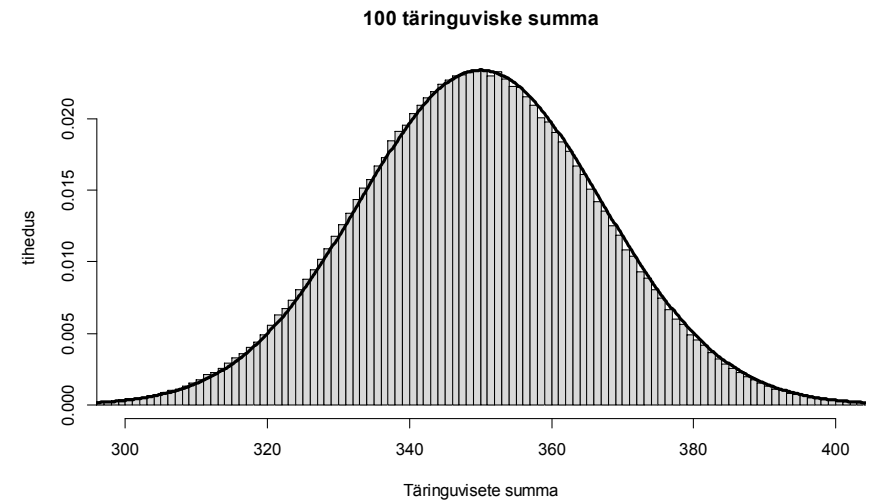
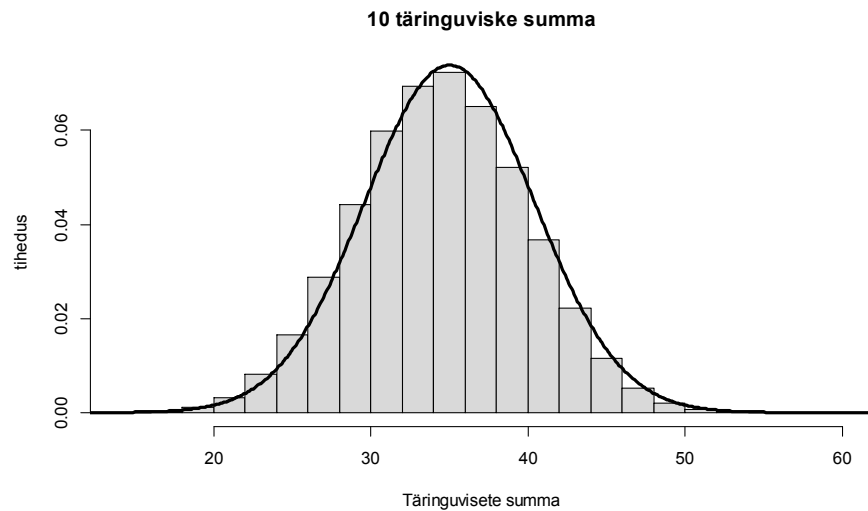
Normaaljaotuse tekkimine

(tsentraalne piirteoreem)

Kui juhuslik suurus Y tekib kui väga paljude juhuslike suuruste summa

$$Y = X_1 + X_2 + \dots$$

ja ükski neist juhuslikest suurustest ei domineeri teiste üle (DY ei tohi olla sama suurusjärguga kui DX_i), siis Y on normaaljaotusega juhuslik suurus



Juhuslike suuruste korrutis...

Kui juhuslik suurus Y on saadud paljude juhuslike suuruste korrutamise teel

$$Y = X_1 \cdot X_2 \cdot X_3 \cdot \dots$$

siis

$$\log(Y) = \log(X_1) + \log(X_2) + \log(X_3) + \dots$$

on normaaljaotusega juhuslik suurus...

Ükskord normaaljaotus...

- Kui juhuslik suurus X on normaaljaotusega, siis $aX+b$ on ka normaaljaotusega juhuslik suurus (a, b – suvalised konstandid)
- Kui X_1 ja X_2 on normaaljaotusega juhuslikud suurused, siis X_1+X_2 on ka normaaljaotusega*

* väide kehtib peaaegu alati (praktiliselt), kuid kunstlikult on võimalik konstrueerida olukord, kus antud väide ei kehti.

Valim ja populatsioon standardviga

Mõisteid

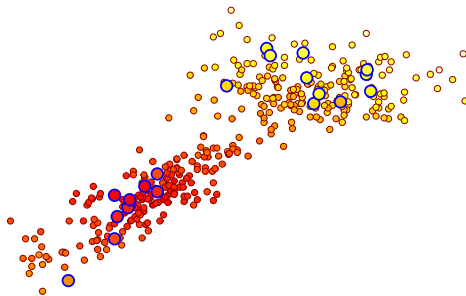
Populatsioon on kõigi objektide, isendite, esemete, nähtuste või seisundite kogum, mille kohta soovitakse järeldusi teha. Sageli kasutatakse populatsiooni asemel ka üldkogumi mõistet. Populatsiooni defineerides piiritletakse ära uuritav objekt (ruumis, ajas, katsetingimuste kaudu,...).

Populatsiooni neid objekte, mida on vaadeldud või uurimiseks välja valitud, kutsutakse *valimiks*.

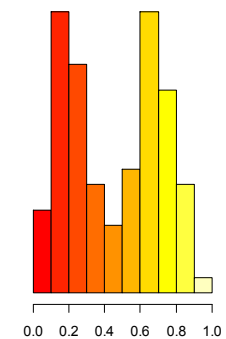
Valimit, kus uuritava tunnuse jaotus on samasugune kui populatsioonis, nimetatakse *esindavaks valimiks*.

14

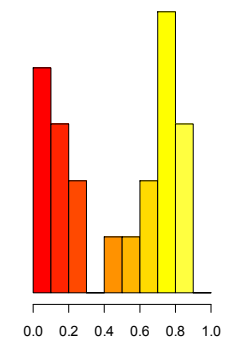
Populatsioon ja juhuslik valim



Populatsioon



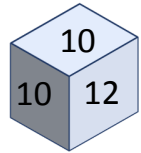
juhuslik valim



Miks töötab?

Populatsioon: 10, 10, 12, 12, 12, 17

Täring:



x	10	12	17
$P(X=x)$	2/6	3/6	1/6

22

Kuidas saada esindavat valimit?

Juhuslik valik

Populatsiooni ja valimi jaotus

Uuritava tunnuse väärtused	Populatsiooni jaotus	Valimi jaotus			
		n=20	n=40	n=100	n=500
Hall	60%	45%	62,5%	62,0%	60,0%
Must	35%	55%	30,0%	36,0%	34,8%
Valge	5%	0%	7,5%	2,0%	5,2%

23

Lihtne juhuslik valik (tagasipanekuga)

1. Kõigil uuritavasse populatsiooni kuuluvatel isenditel/objektidel on võrdne tõenäosus sattuda meie valimi 1. vaatluseks (elemendiks);
2. Kõigil uuritavasse populatsiooni kuuluvatel isenditel/objektidel on võrdne tõenäosus sattuda meie valimi 2. elemendiks (sõltumata sellest, keda või mida me oleme juba vaadelnud);
3.
-
-
- n. Kõigil uuritavasse populatsiooni kuuluvatel isenditel/objektidel on võrdne tõenäosus sattuda meie valimi n . elemendiks (sõltumata sellest, keda või mida me oleme juba vaadelnud);

24

populatsioon

valim

keskväärtus (EX, μ)

keskmine ($\bar{x}$)

pop. dispersioon

valimi dispersioon (s^2)

(DX, σ^2)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

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

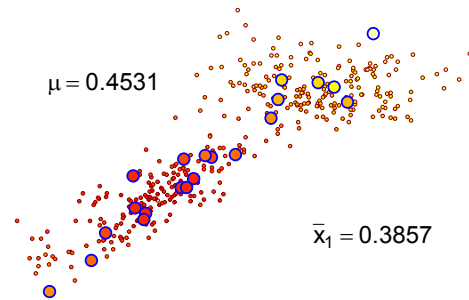
pop. mediaan

valimi mediaan

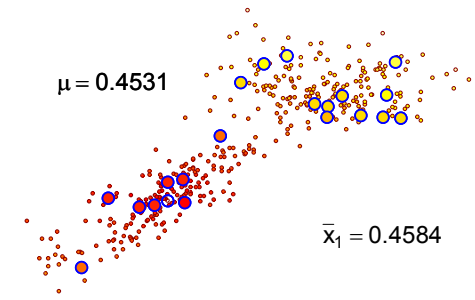
....

....

Hinnangu täpsus valim 1



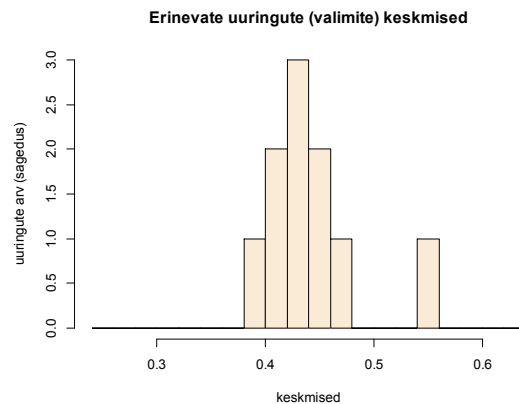
Hinnangu täpsus valim 2



Valimikeskmiste standardhälve - standardviga

valim keskmine

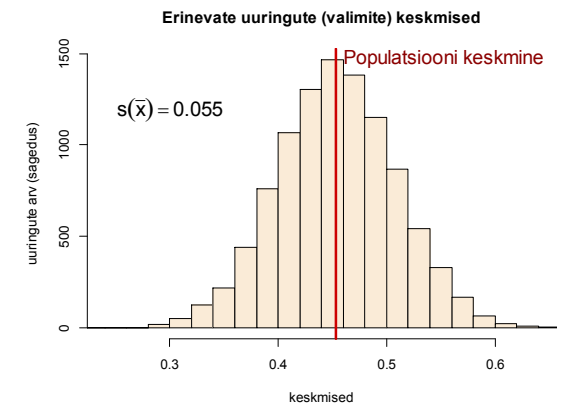
1	0.3857
2	0.4584
3	0.4650
4	0.4349
5	0.5468
6	0.4121
7	0.4459
8	0.4123
9	0.4347
10	0.4393



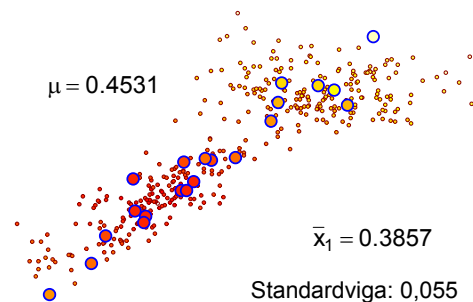
Valimikeskmiste standardhälve - standardviga

valim keskmine

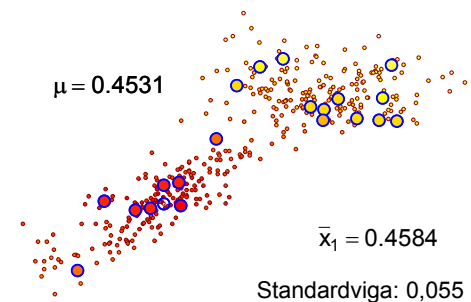
1	0.3857
2	0.4584
3	0.4650
4	0.4349
5	0.5468
6	0.4121
7	0.4459
8	0.4123
9	0.4347
10	0.4393
..	



Hinnangu täpsus valim 1



Hinnangu täpsus valim 2



Matemaatiline trikk

$$\begin{aligned}
 D(\bar{x}) &= D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \xleftarrow{D(cX) = c^2 D(X)} \\
 &= \frac{1}{n^2} D\left(\sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n D(X_i) \xleftarrow{\text{Kui vaatlused on sõltumatud} \\ &\quad \text{– see, kes järgmisena} \\ &\quad \text{valimisse sattub, ei tohi} \\ &\quad \text{sõltuda sellest, kes juba} \\ &\quad \text{valimis ees on.}} \\
 &= \frac{1}{n^2} n \cdot D(X) \xleftarrow{D(X+Y) = D(X) + D(Y), \text{ kui } X \\ &\quad \text{ja } Y \text{ on sõltumatud}} \\
 &= \frac{D(X)}{n} \xleftarrow{\text{Kui kõik vaatlused on samast} \\ &\quad \text{populatsioonist – sellisest, mis} \\ &\quad \text{vahepeal pole ära muutunud (näiteks} \\ &\quad \text{tänu sellele, et osad vaatlused on} \\ &\quad \text{valimisse "eemaldatud")}}
 \end{aligned}$$

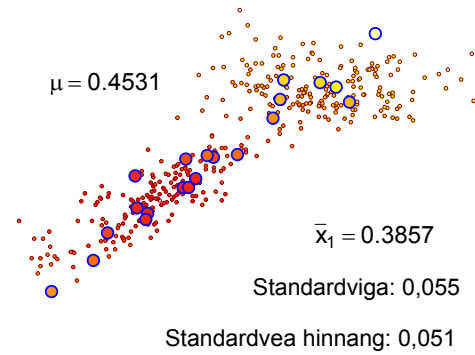
Matemaatiline trikk

$$\begin{aligned}
 D(\bar{x}) &= D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} D\left(\sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n D(X_i) \\
 &= \frac{1}{n^2} n \cdot D(X) \\
 &= \frac{D(X)}{n}
 \end{aligned}$$

Standardviga

$$\begin{aligned}
 \sigma(\bar{x}) &= \sqrt{D(\bar{x})} \\
 &= \sqrt{D(X)/n} \\
 &\approx \sqrt{s^2/n} \\
 &= s/\sqrt{n}
 \end{aligned}$$

Hinnangu täpsus valim 1



Hinnangu täpsus valim 2

