

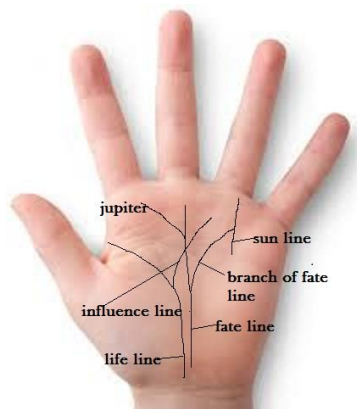
REGRESSIOON- ANALÜÜS I

TEEMAD

- Matemaatiline mudel
- Regressioonmudel
- Vähimruutude meetod
- Regressioonmudeli mittepööratavus
- Seose tugevus ja determinatsioonikordaja
- Parameetrite standardvead ja usalduspiirid
- Prognoosi usalduspiirid
- Mittelineaarsed mudelid
- Jääkide analüüs

ENNUSTAMINE

Mida toob tulevik?
Kuidas ennustada?



Ainult matemaatiliste mudelite kasutamine võimaldab teha paikapidavaid ennustusi.

MATEMAATILINE MUDEL

Mudel on reaalses maailmas esineva objekti analoog, mis asendab seda objekti tunnetusprotsessis.

Mudel peab

- välja tooma olulise;
- jätma kõrvale mitteolulise.

Matemaatiline mudel on mingit reaalses maailmas eksisteerivat nähtust kirjeldavate matemaatiliste seoste kogum.

$$y = f(x_1, x_2, \dots, a, b, \dots)$$

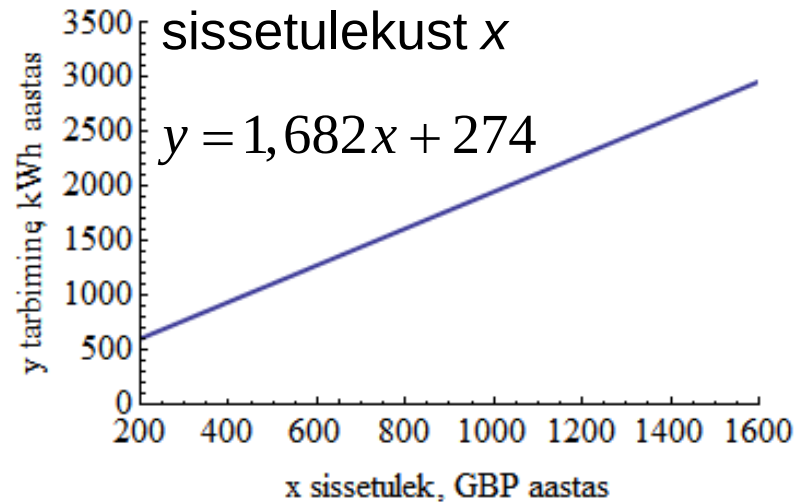
y sõltuv muutuja

$x_1, x_2, \dots$ sõltumatud muutujad ehk argumendid

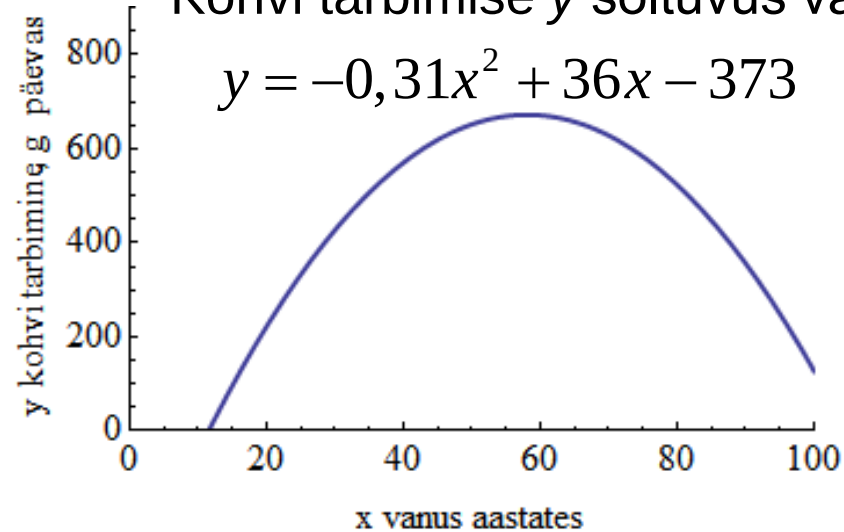
$a, b, \dots$ mudeli parameetrid (konstandid)

NÄIDE: ERINEVAD MATEMAATILISED MUDELID

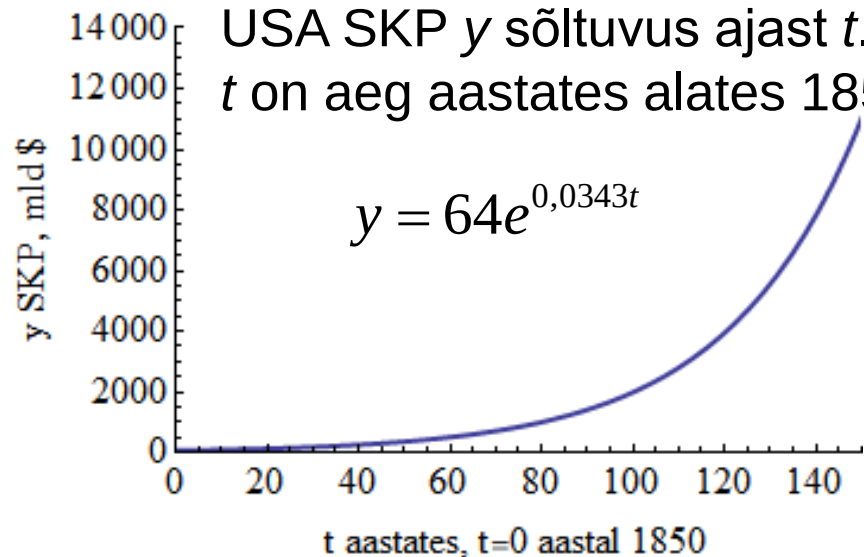
Elektrienergia tarbimise y sõltuvus
sissetulekust x



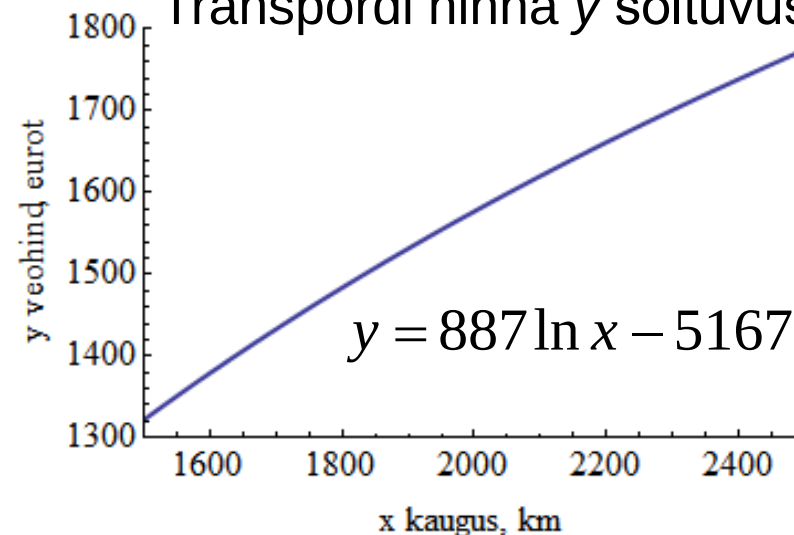
Kohvi tarbimise y sõltuvus vanusest x



USA SKP y sõltuvus ajast t .
 t on aeg aastates alates 1850.

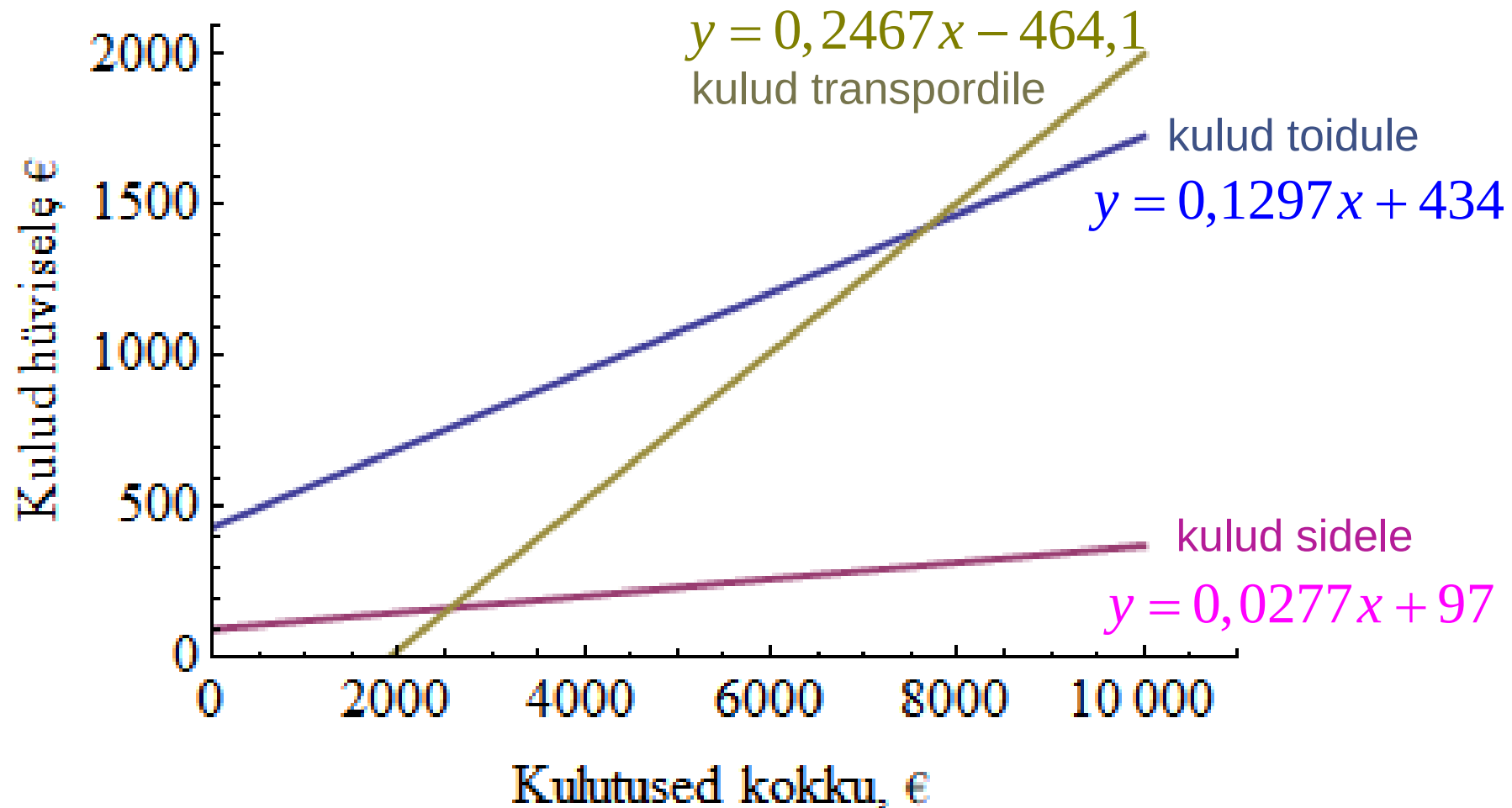


Transpordi hinna y sõltuvus kaugusest x



NÄIDE: ERINEVAD LINEAARSED MUDELID

Kulud erinevatele hüvistele y ja kogukulud x pereliikme kohta aastas (2012.a.)



LINEAARSE MUDELI ÜLDKUJU JA PARAMEETRITE TÕLGENDUS

Üldkuju $y = ax + b$

Lineaarliikme kordaja a näitab, kui palju muutub y , kui x suureneb 1 võrra.

$$y_1 = ax_1 + b$$

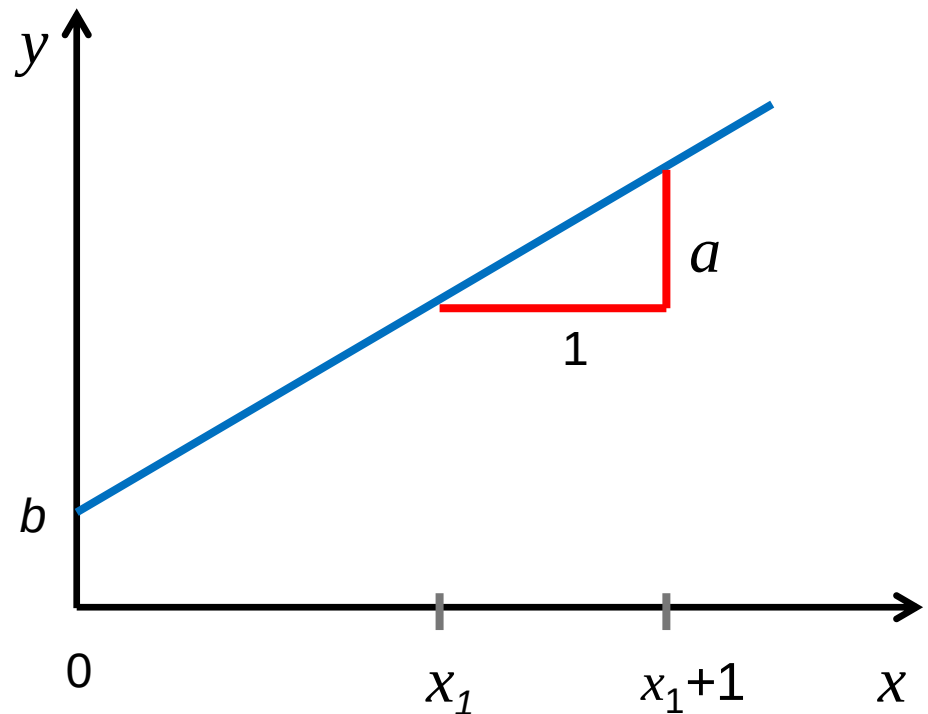
$$y_2 = a(x_1 + 1) + b =$$

$$= ax_1 + a + b =$$

$$= ax_1 + b + a = y_1 + a$$

Vabaliige b näitab sõltuva muutuja y väärtust, kui $x=0$.

$$y = ax + b, \quad x = 0 \Rightarrow y = b$$



Demo: lineaarse mudeli parameetrid

LINEAARSE MUDELI KONKREETNE KUJU

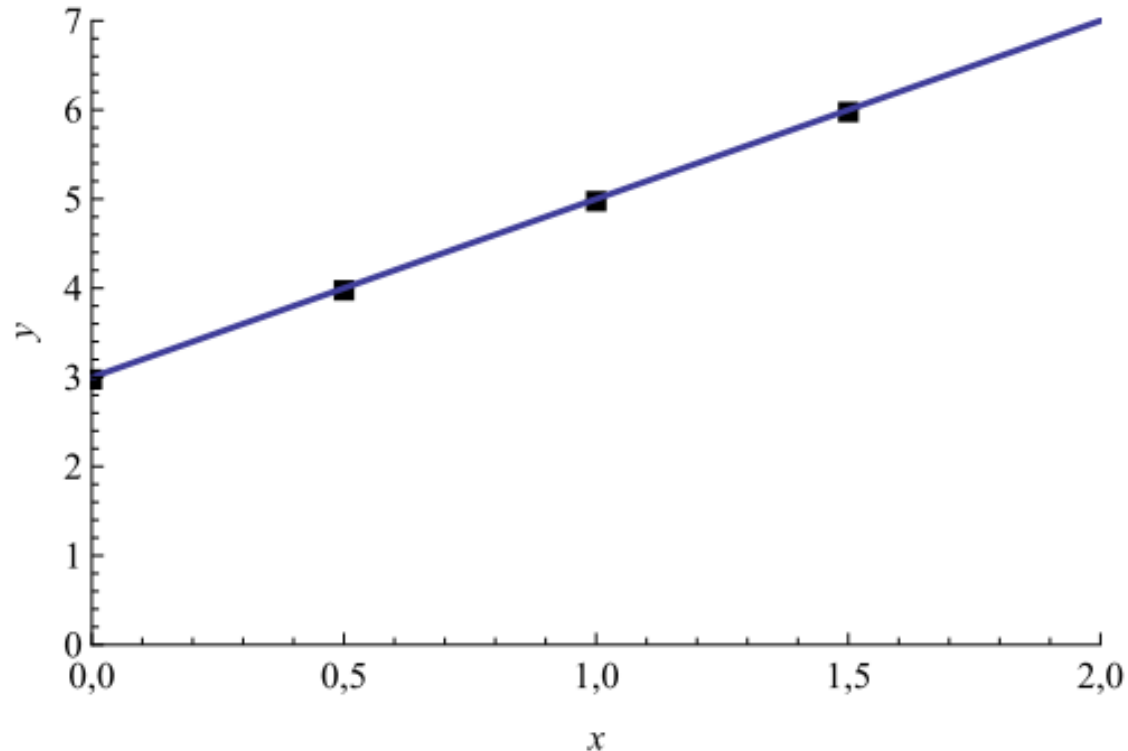
Üldkuju $y = ax + b$



Anname parameetritele väärtused $a = 2$, $b = 3$

Konkreetne kuju

$$y = 2x + 3$$



?

EMPIIRILISED ANDMED => MUDEL

On olemas

1) empiirilised andmed

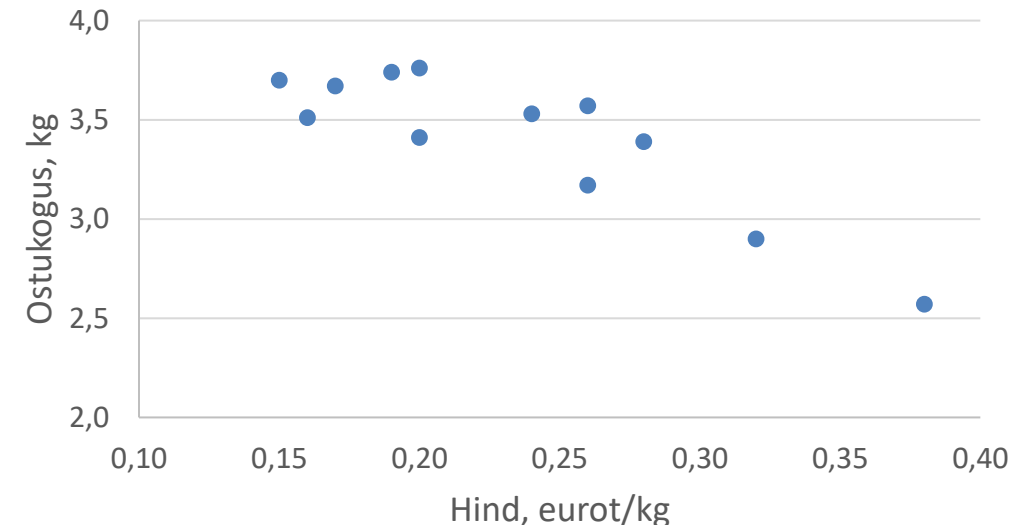
Kartuli hind ja ostukogus
pereliikme kohta kuus
aastatel 1996-2007.

Hind x, €/kg	Ostukogus y, kg
0,17	3,67
0,16	3,51
0,20	3,41
0,20	3,76
0,15	3,70
...	

2) mudeli üldkuju $y = ax + b$

Kuidas empiiriliste andmete põhjal
määrata mudeli parameetrid a ja b ?

Regressioonanalüüs!



Allikas: Eesti Statistikaameti andmebaas, tabel LE13, toidukaupade
keskmine ostukogus ja hind leibkonnaliikme kohta kuus.

NÄIDE: PIKKUSE MODELLEERIMINE

Mingi Eesti mehe pikkus (cm)

$$\text{PIKKUS} = 179 + \varepsilon$$

Eesti meeste keskmine pikkus

Juhuslik komponent

2 – 16 aastase poisslapse pikkus (cm)

$$\text{PIKKUS} = 80,4 + 6 \cdot \text{VANUS} + \varepsilon$$

Keskmine pikkus sõltub vanusest

Juhuslik komponent

REGRESSIOONMUDEL

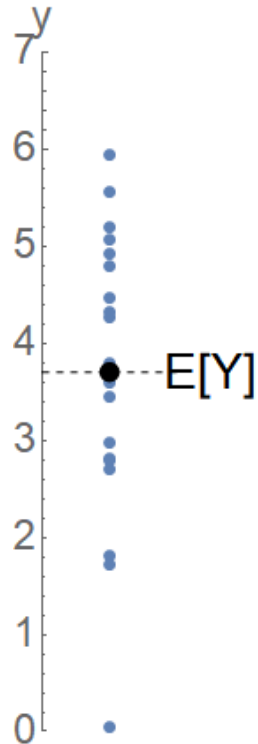
$$y_i = \text{deterministlik komponent} + \text{juhuslik komponent}$$

Deterministlik komponent on see oluline osa, mille mudel peab välja tooma.

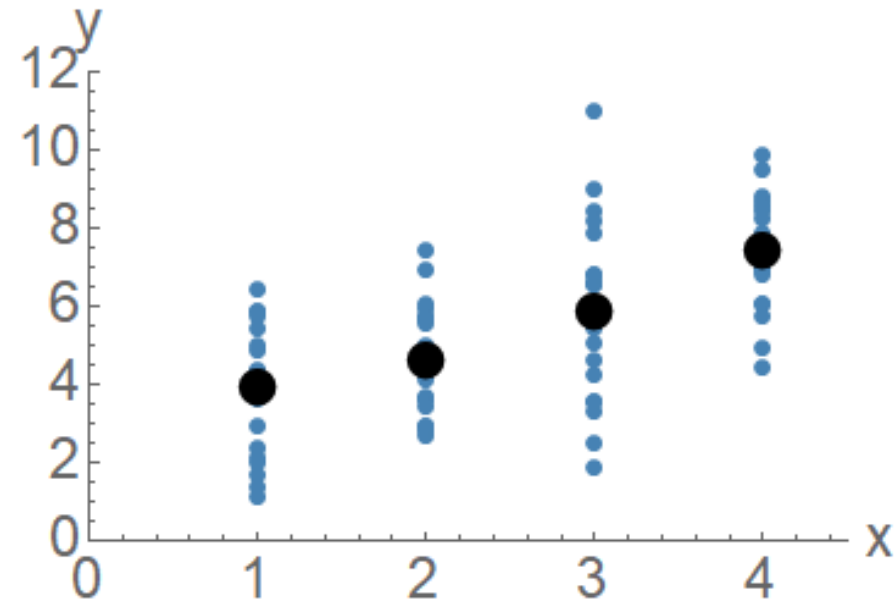
Regressioonanalüüsi käigus leitakse regressioonmudeli deterministlik komponent.

Deterministlik komponent = tinglik keskväärtus $E[Y|X]$

DETERMINISTLIK KOMPONENT = TINGLIK KESKVÄÄRTUS



Tunnus Y ei sõltu teisest tunnusest.
Väärtused hajuvad juhuslikult.
Nende keskväärtus on $E[Y]$.



Tunnus Y sõltub tunnusest X .
Iga X väärtuse korral on tunnuse Y
keskväärtus erinev.
See on tinglik keskväärtus $E[Y|X]$.

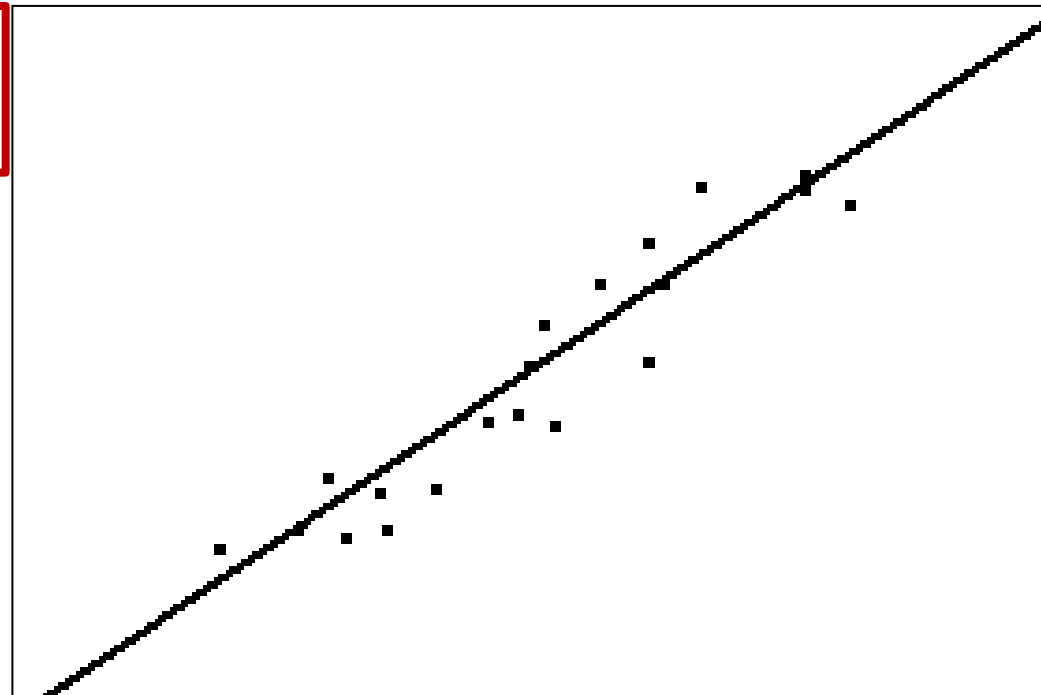
LINEAARNE REGRESSIOONMUDEL

Üldkuju

$$y = \alpha X + \beta + \varepsilon$$

mudeli parameetrid

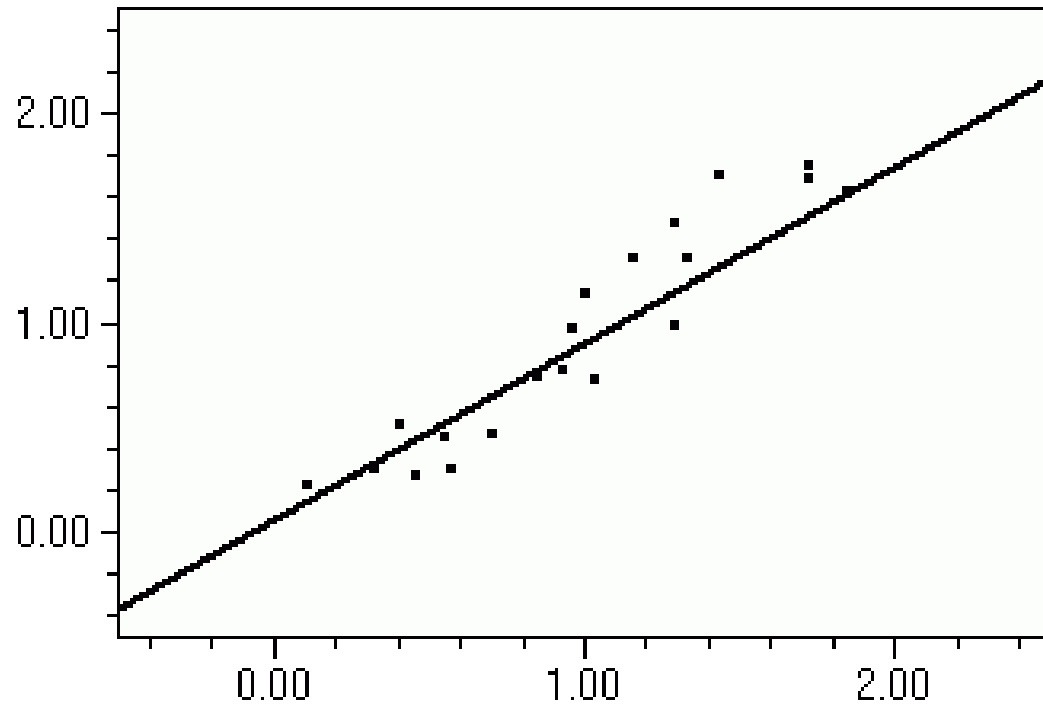
Regressioonanalüüsi abil tuleb hinnata nende parameetrite arväärtusi.



Lineaarse mudeli hindamine = parima sirge leidmine

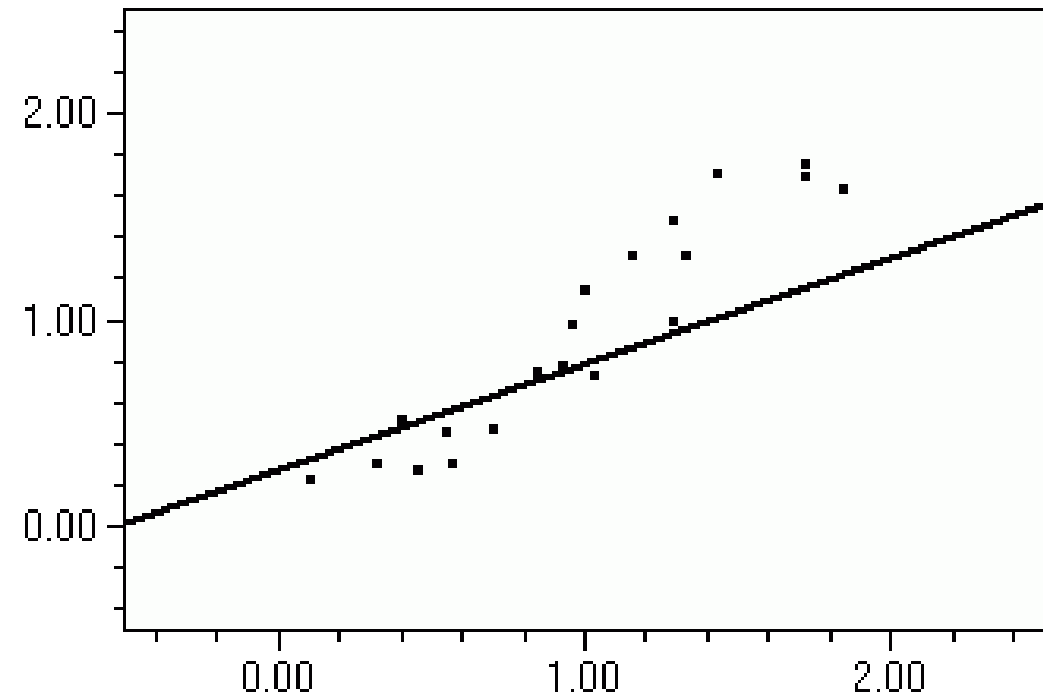
KUMB SIRGE LÄBIB PUNKTIPARVE PAREMINI?

A



$$y = 0,81x + 0,10$$

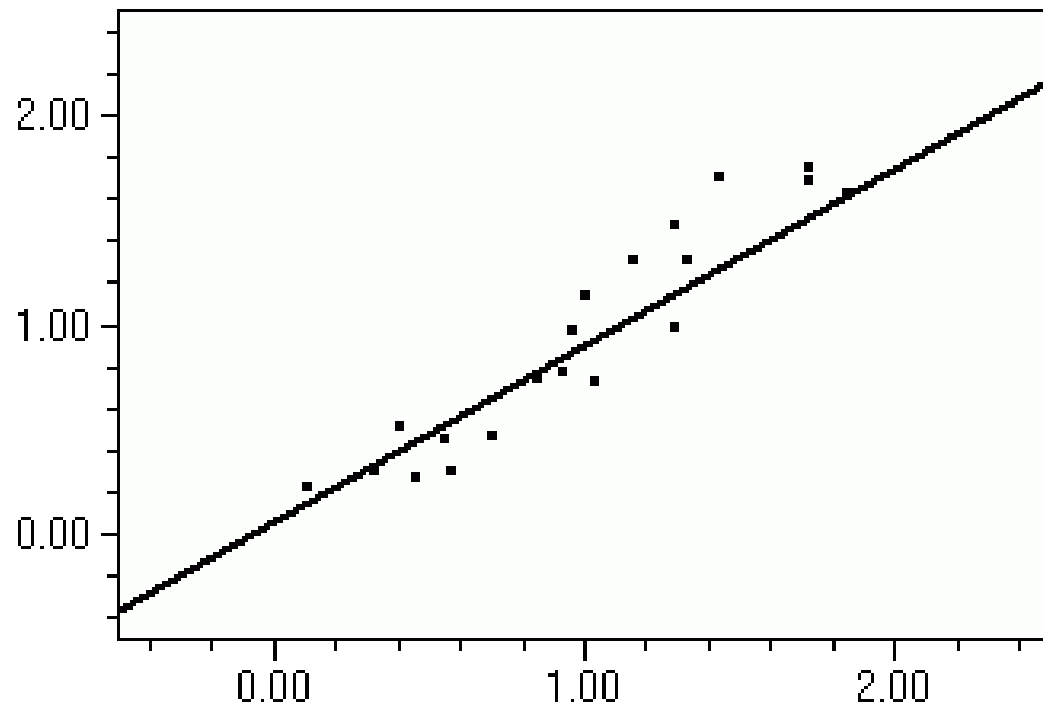
B



$$y = 0,51x + 0,28$$

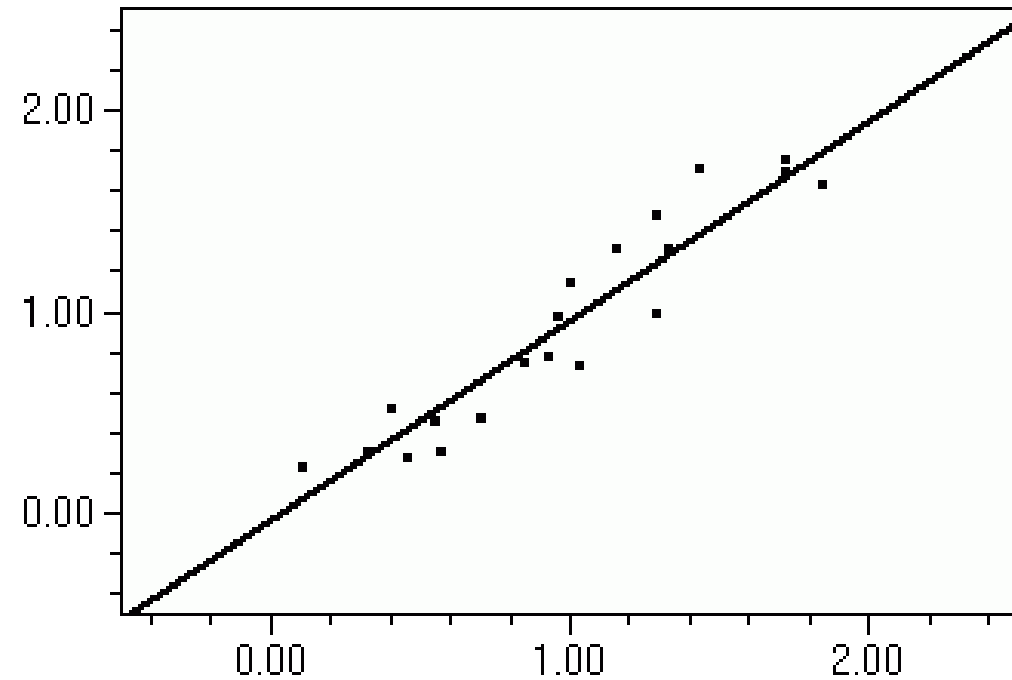
KUMB SIRGE LÄBIB PUNKTIPARVE PAREMINI?

A



$$y = 0,81x + 0,10$$

B



$$y = 0,99x - 0,04$$

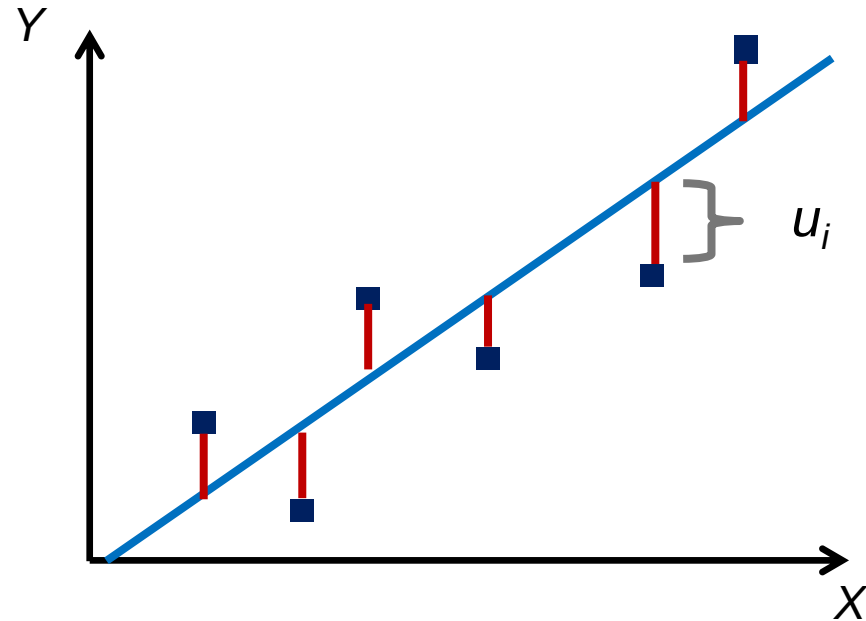
Vaja objektiivset kriteeriumi!

VÄHIMRUUTUDE MEETOD

Objektiivne kriteerium:

Empiiriliste punktide ja sirge vastavate punktide vaheliste kauguste ruutude summa on minimaalne.

$$\sum_i u_i^2 \rightarrow \min$$

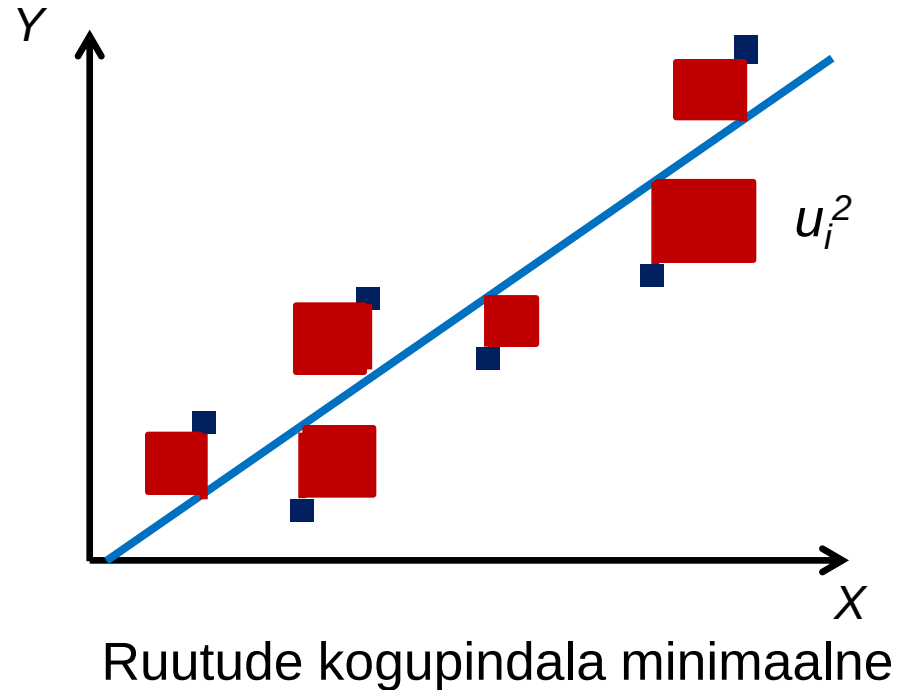


VÄHIMRUUTUDE MEETOD

Objektiivne kriteerium:

Empiiriliste punktide ja sirge vastavate punktide vaheliste kauguste ruutude summa on minimaalne.

$$\sum_i u_i^2 \rightarrow \min$$



VALEMID MUDELI PARAMEETRITE JAOKS, TULETAMINE

$$y_i = ax_i + b + u_i$$

$$\sum_{i=1}^n u_i^2 \rightarrow \min$$

$$u_i = y_i - ax_i - b$$

$$\sum_{i=1}^n (y_i - ax_i - b)^2 \rightarrow \min$$

$$a, b = ?$$

Kahe muutuja funktsiooni minimeerimine.

Tähistame $S(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$

1. Leida S osatuletised a ja b suhtes.
2. Lahendada võrrandsüsteem.

$$\begin{cases} \frac{\partial S(a, b)}{\partial a} = 0 \\ \frac{\partial S(a, b)}{\partial b} = 0 \end{cases}$$

VALEMID MUDELI PARAMEEETRITE JAOKS

Mudeli üldkuju $y_i = \alpha X_i + \beta + \varepsilon_i$

Parameetrite hinnangud, leitud osatuletistega võrrandsüsteemi lahendamisel.

Parameetri α hinnang $a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Parameetri β hinnang $b = \bar{y} - a\bar{x}$

NÄIDE: KARTULI HIND JA OSTUKOGUS

Mudel $y = 4,48 - 4,59x + \varepsilon$

y kartuli ostukogus pereliikme kohta kuus, kg

x kartuli hind, €/kg

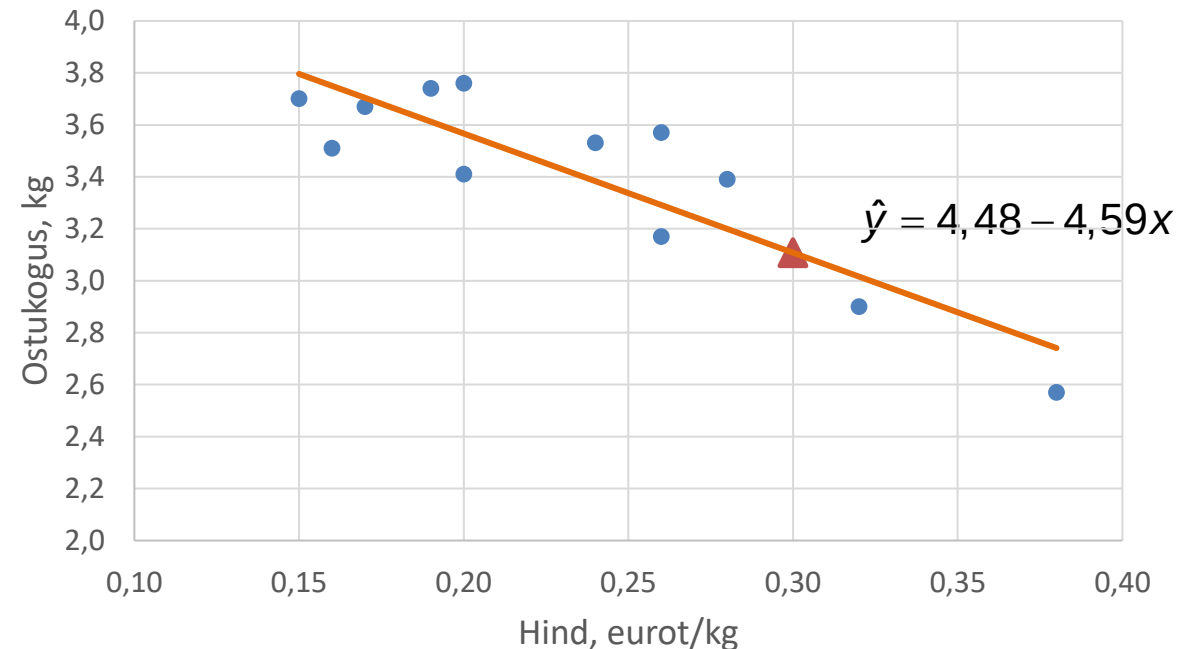
Deterministlik komponent ehk

mudelväärtus:

$$\hat{y} = 4,48 - 4,59x$$

Kui kartuli kg hind on 0,3 eurot,
milline oleks keskmine ostukogus
pereliikme kohta kuus?

$$\hat{y} = 4,48 - 4,59 \cdot 0,3 \approx 3,1 \text{ kg kuus}$$



NÄIDE: KARTULI HIND JA OSTUKOGUS

Mudel $y = 4,48 - 4,59x + \varepsilon$

y kartuli ostukogus pereliikme kohta kuus, kg

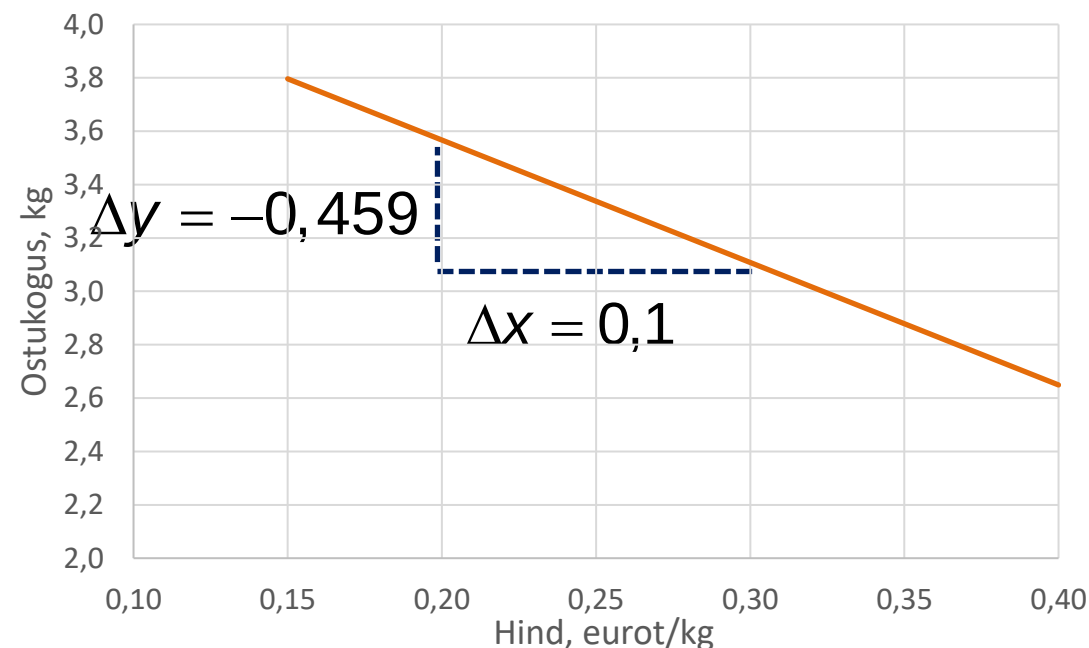
x kartuli hind, €/kg

KORDAJA TÕLGENDAMINE

Kui kg hind suureneb 1 euro võrra, siis prognoositav kartuli ostukogus pereliikme kohta väheneb 4,59 kg võrra kuus.

Ehk

kui hind suureneb 0,1 euro ehk 10 senti võrra, siis prognoositav kartuli ostukogus pereliikme kohta väheneb 0,459 kg võrra kuus.



NÄIDE: KOGUKULUD JA KULUD TOIDULE

$$\hat{y}_T = 434 + 0,1297x$$

x kogukulud, € pereliikme kohta aastas

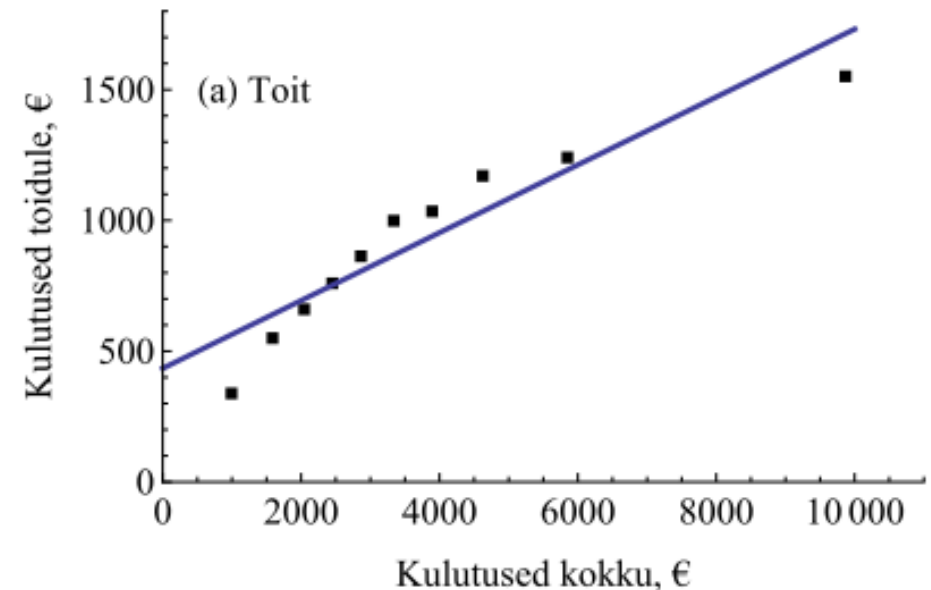
y kulud toidule, € pereliikme kohta aastas

Tõlgendame mudelit

Kogukulude suurenedes 1 € võrra suurenesid kulutused toidule 0,1297 euro võrra. Tarbimise piirmäär.

Kogukulude suurenedes 1000 € võrra suurenesid kulutused toidule 129,7 € võrra.

Igast täiendavalt kulutatud 1000 eurost suunati 129,7 € täiendavalt toiduainete tarbimisele.

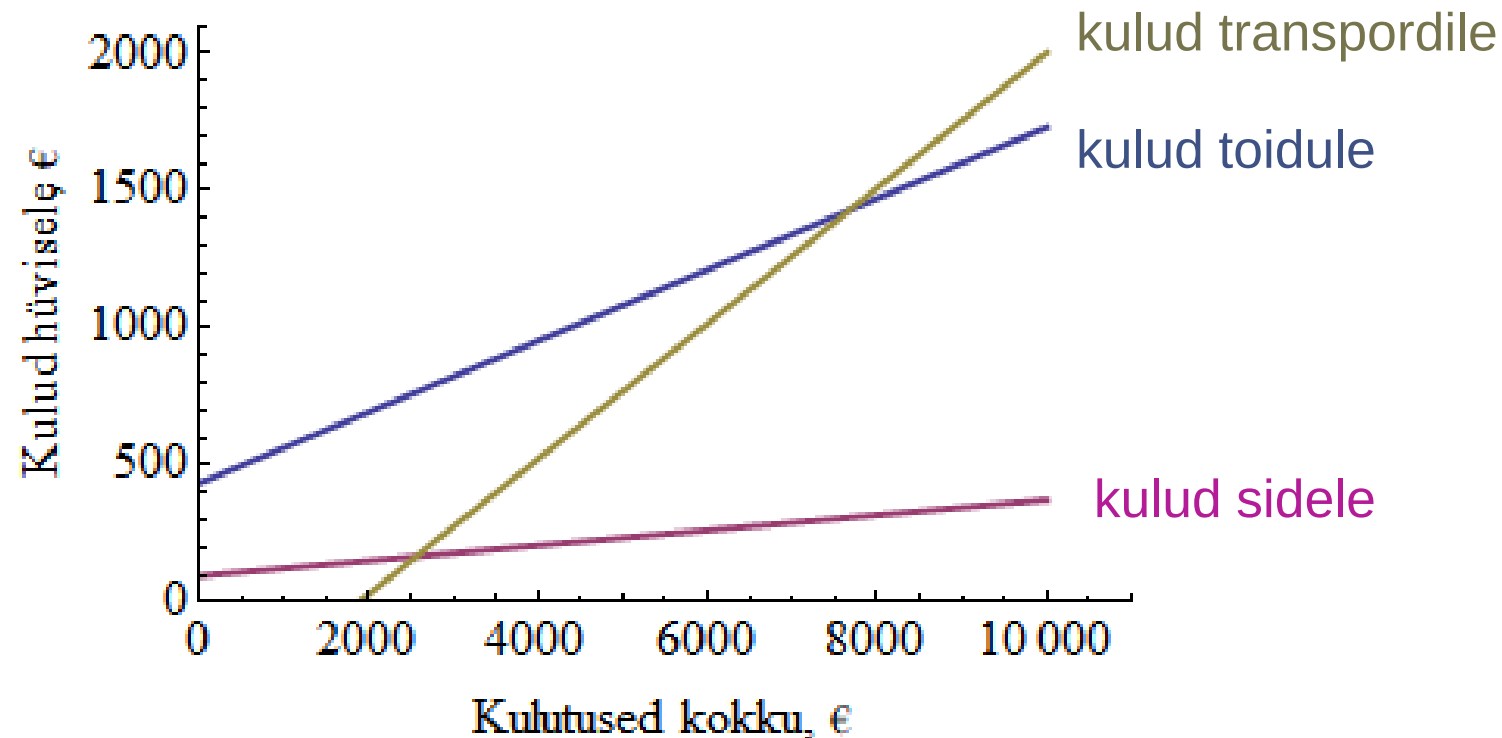


NÄIDE: TARBIMISFUNKTSIOONIDE VÕRDLU

Kulud toidule $\hat{y}_T = 434 + 0,1297x$

Kulud sidele $\hat{y}_S = 97 + 0,0277x$

Kulud transpordile $\hat{y}_{TR} = -464,1 + 0,2467x$



REGRESSIOONI MITTEPÖÖRATAVUS

Korrelatsioonikordaja on sümmeetriline x ja y suhtes

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$

Regressioonmudeli parameetrid EI OLE sümmeetrilised x ja y suhtes

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

Regressioonanalüüs ei ole pööratav.

KUMB ON X, KUMB Y

$$y = ax + b$$

x argumenttunnus, seletav tunnus, sõltumatu tunnus, regressor
 y funktsioontunnus, sõltuv tunnus

Tuleb õigesti määrata, kumb on x , kumb y .

Kinnisvaratehingud:

maja hind y

maja pindala x

telesaate vaadatavus x

reklaamiminuti hind y

SKP elaniku kohta x

telerite arv 1000 el. kohta y

SKP y

tööjõud x

NÄIDE STATISTIKA EKSAMILT

Andmed

Pere	Pereliikmete arv	Kulud toidule aastas, eurot
1	2	746
2	5	1470
...	...	...

Ülesanne.

Leida mudel, mis kirjeldab toidule tehtavate kulude sõltuvust pereliikmete arvust.

Üliõpilane leidis mudeli

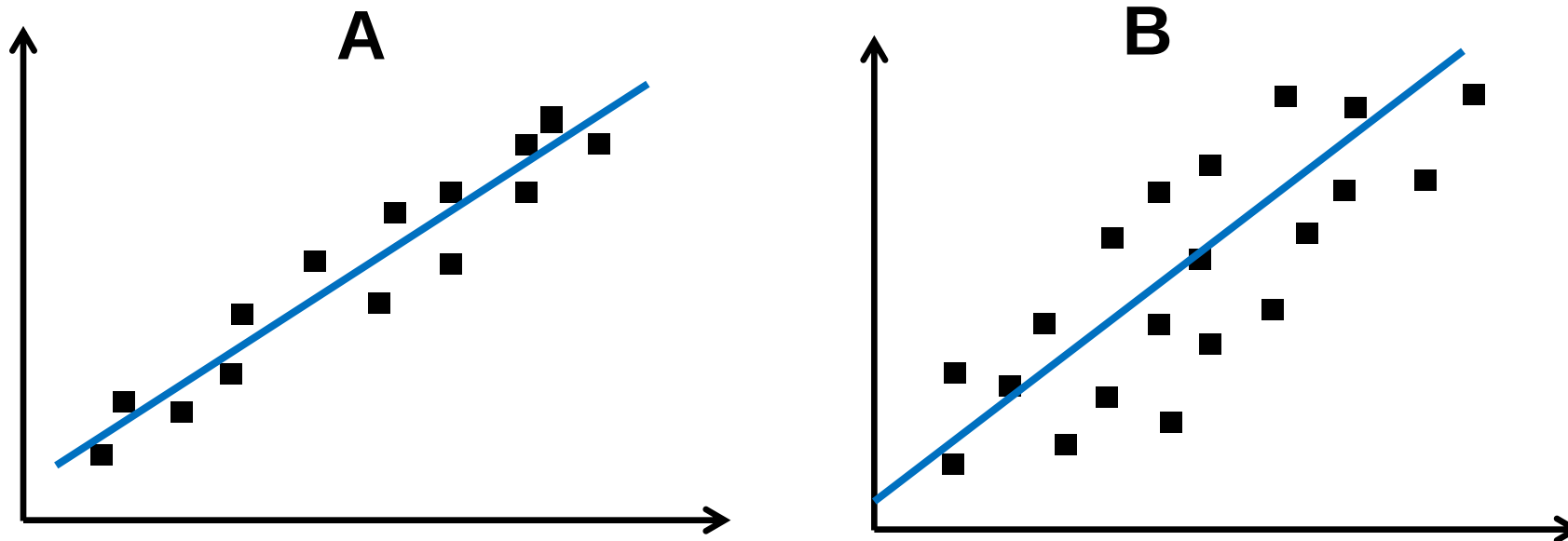
$$\hat{y} = 2,4 + 0,0003x$$



Pereliikmete arv

Kulud toidule

SEOSE TUGEVUS



Kumb sirge kirjeldab oma punktisarve paremini?

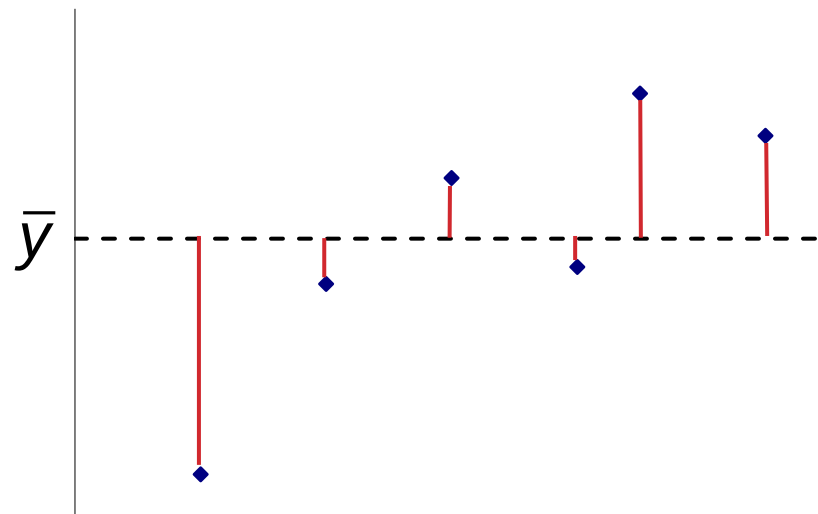
Vaja objektiivset karakteristikut!

KOGUHAJUVUS

Sõltuva tunnuse Y
koguhajuvus

$$SST = \sum (y_i - \bar{y})^2$$

Total Sum of Squares



Koguhajuvus = regressioonhajuvus + jääkhajuvus

$$SST = SSR + SSE$$

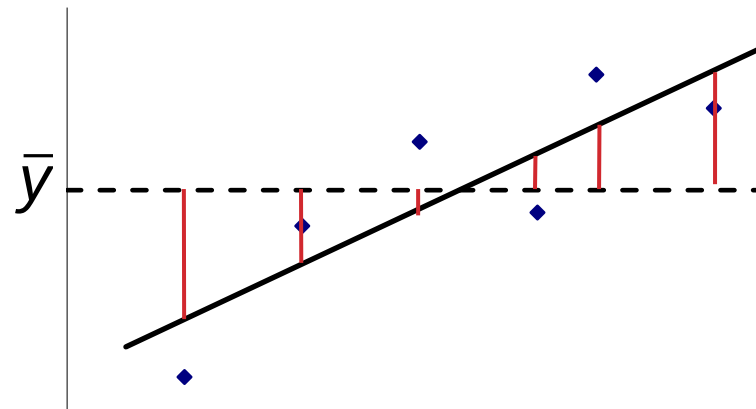
$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y}_i)^2 + \sum (y_i - \hat{y}_i)^2$$

REGRESSIOONHAJUVUS, JÄÄKHAJUVUS

Regressioonhajuvus $SSR = \sum (\hat{y}_i - \bar{y})^2$

Sum of Squares due to Regression

Ka seletatud hajuvus.

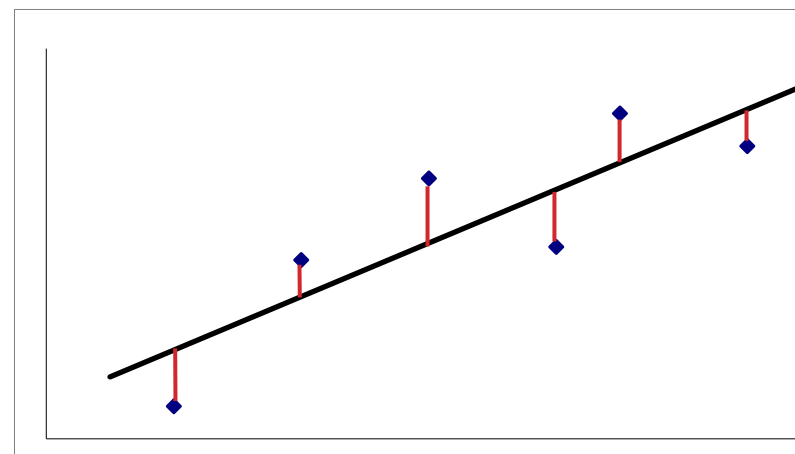


Jääkhajuvus

$$SSE = \sum (y_i - \hat{y}_i)^2$$

Sum of Squared Errors

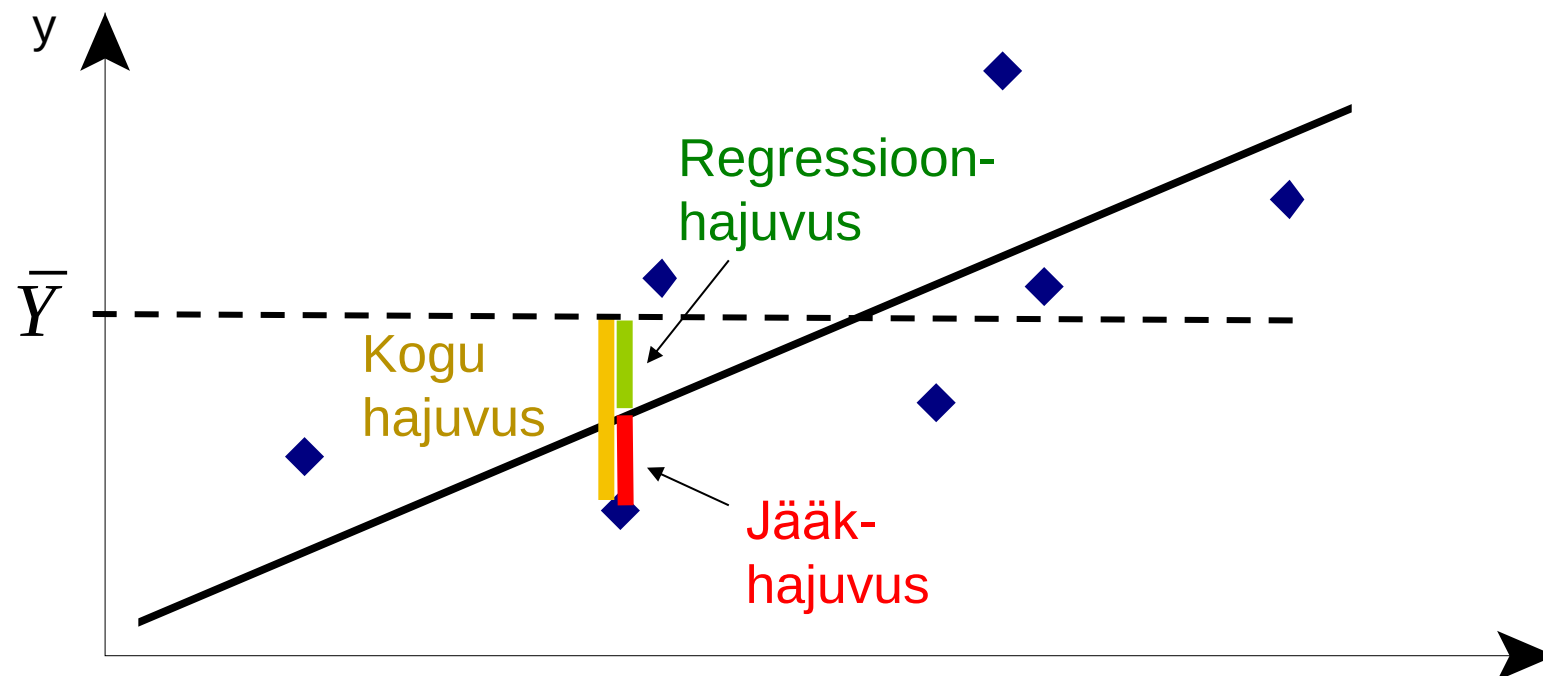
Ka seletamata hajuvus.



Koguhajuvus = regressioonhajuvus + jääkhajuvus

DETERMINATSIOONIKORDAJA R^2

Arvutamise
põhimõte:



$$R^2 = \frac{\text{regressioonhajuvus}}{\text{koguhajuvus}} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Näitab, kui suur osa koguhajumisest on mudeli poolt ära seletatud.

DETERMINATSIOONIKORDAJA R^2

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$0 \leq R^2 \leq 1$$

- **Maksimaalne** väärtus 1, kui jääkhajuvus $SSE=0$.
St. jääkhajuvust ei esine, kõik on mudeli poolt ära seletatud.
- **Minimaalne** väärtus 0, kui regressioonhajuvus $SSR=0$.
St. mudeliga pole midagi ära seletatud.
- Ühikuta suurus.

NÄIDE: KARTULI OSTUKOGUS JA HIND

Kartuli ostukoguse y (kg) sõltuvus hinnast x

$$\hat{y} = 4,48 - 4,59x \quad R^2 = 0,77$$

77% kartuli tarbimise varieerumisest on tingitud sellest, et hind muutub.

Determinatsioonikordaja iseloomustab mudeli kirjeldusvõimet.

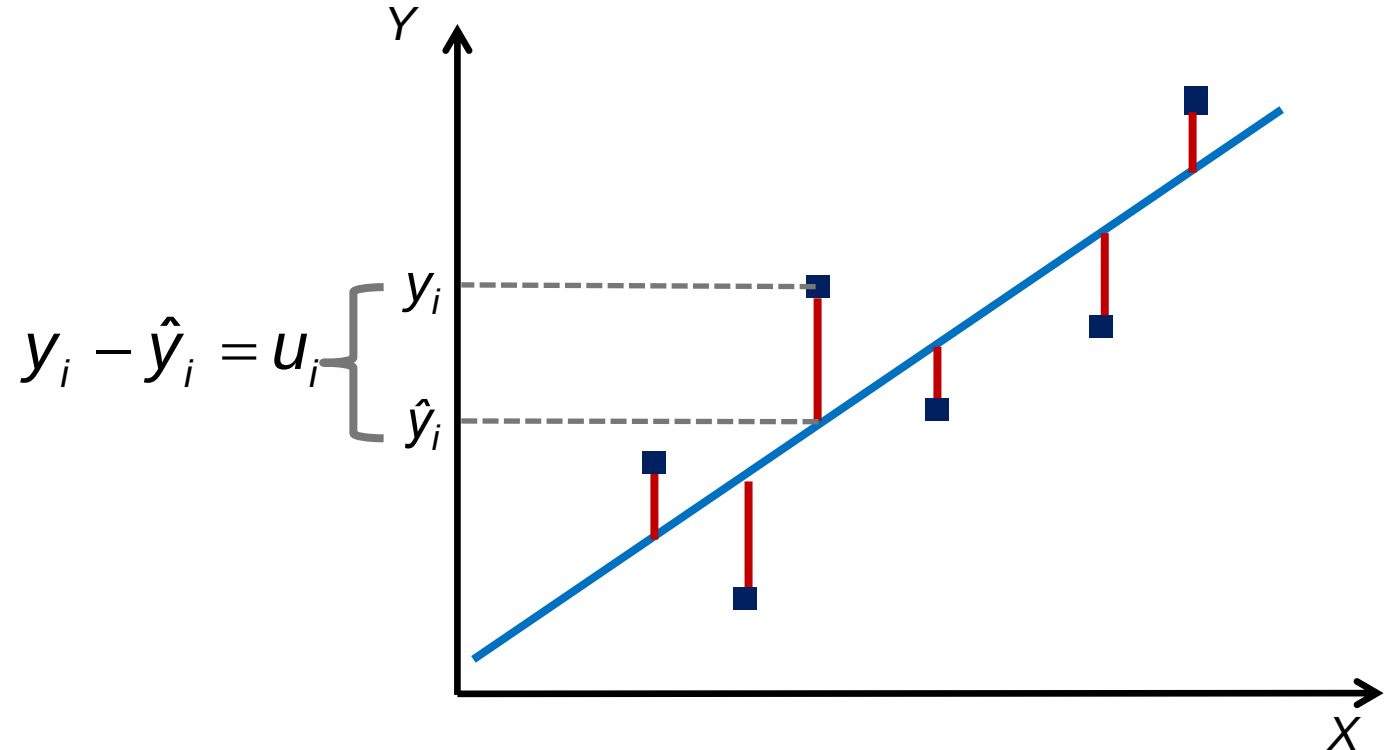
MUDELI STANDARDVIGA

standard error

Standardviga se iseloomustab funktsioontunnuse väärtuste y_i kõrvalekallet mudelväärtustest $\hat{y}_i$.

$$se = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}}$$

$$se = \frac{1}{\sqrt{n - 2}} \sqrt{\sum_{i=1}^n u_i^2}$$



Vähimruutude meetod tähendab mudeli standardvea minimeerimist.

PARAMEETRITE STANDARDVEAD

Mudeli standardvea se alusel leitakse parameetrite standardvead.

Demo:
standardvead

Sirge tõusu a
standardviga

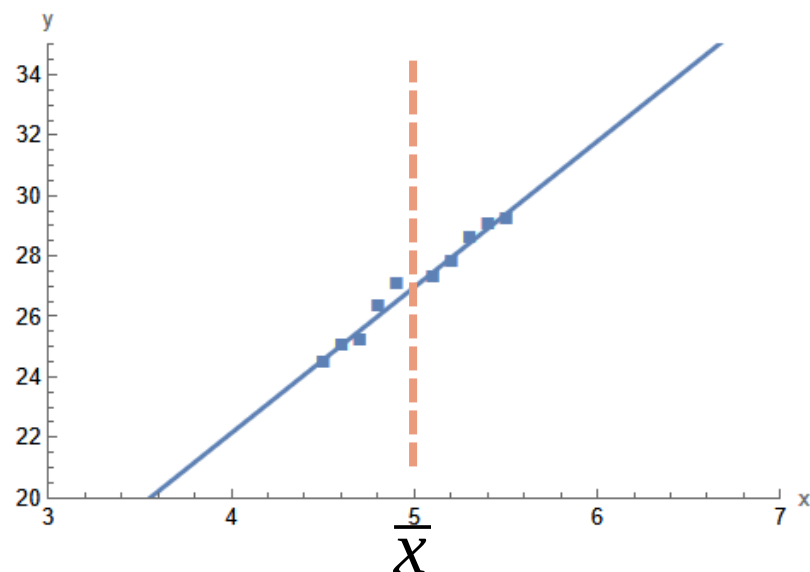
$$se(a) = \frac{se}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Vabaliikme b
standardviga

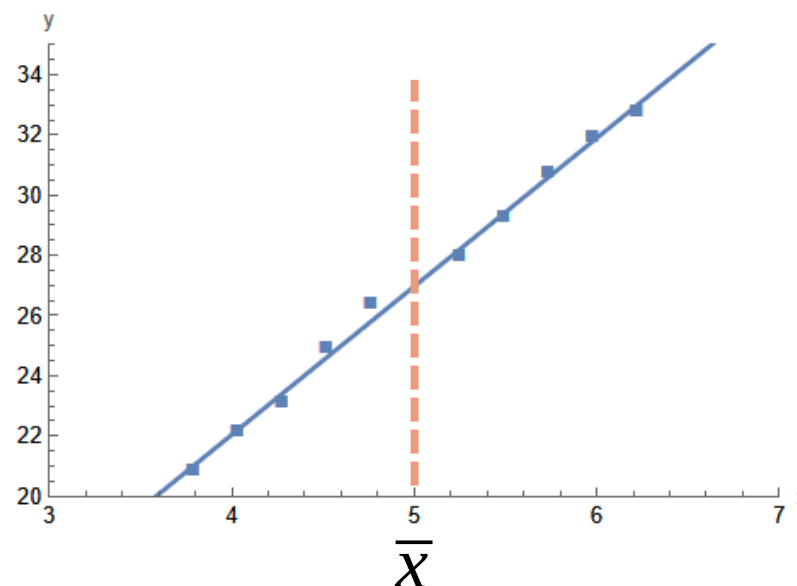
$$se(b) = se \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$$

$\sum (x_i - \bar{x})^2$ iseloomustab x väärtuste hajumist ümber keskmise $\bar{x}$

Hajumine väike: standardvead suured



Hajumine suurem: standardvead väiksemad



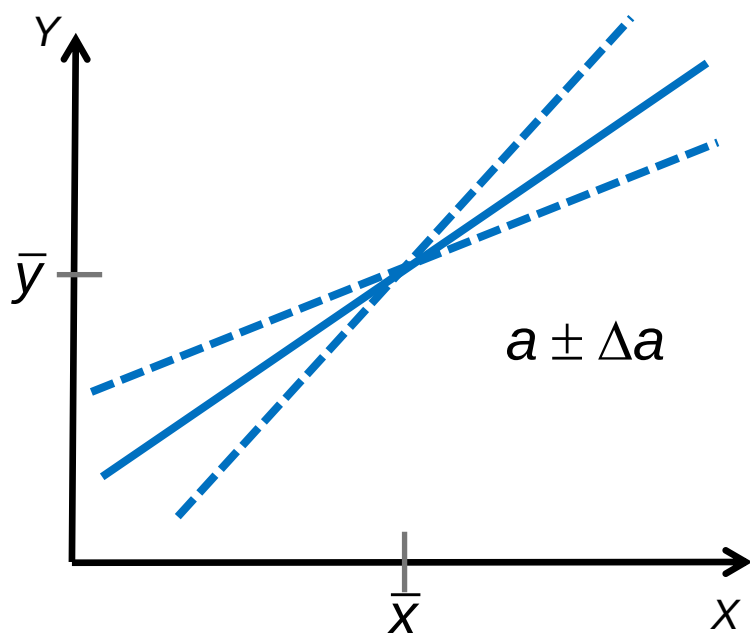
PARAMEETRITE USALDUSVAHEMIKUD

Standardvigade alusel leitakse parameetrite usaldusvahemikud etteantud usaldatavusega (tavaliselt 95%).

Sirge tõus a

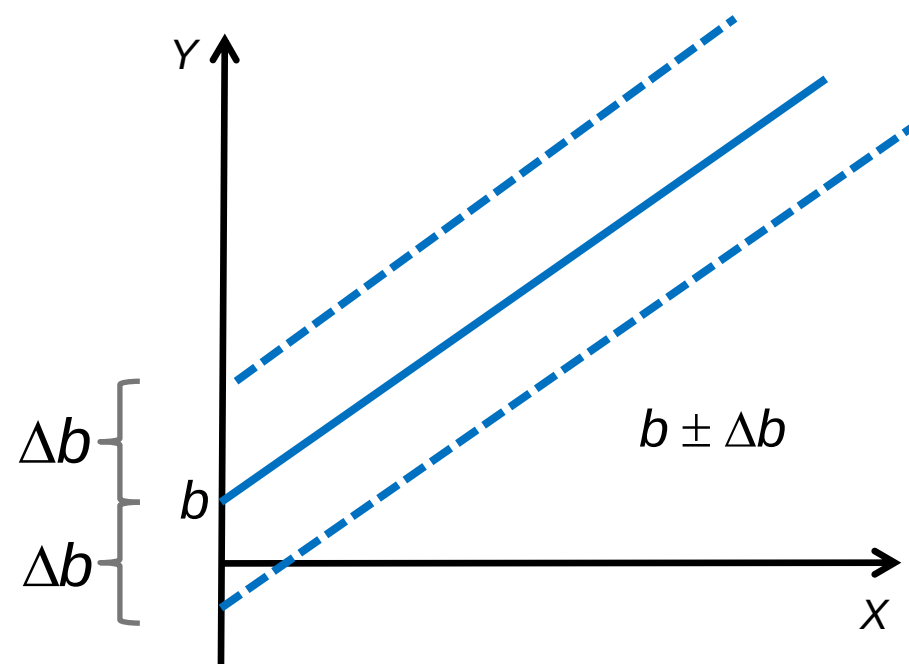
$$\Delta a = t_{\alpha/2}(\nu) se(a)$$

Usaldusvahemiku
poollaius

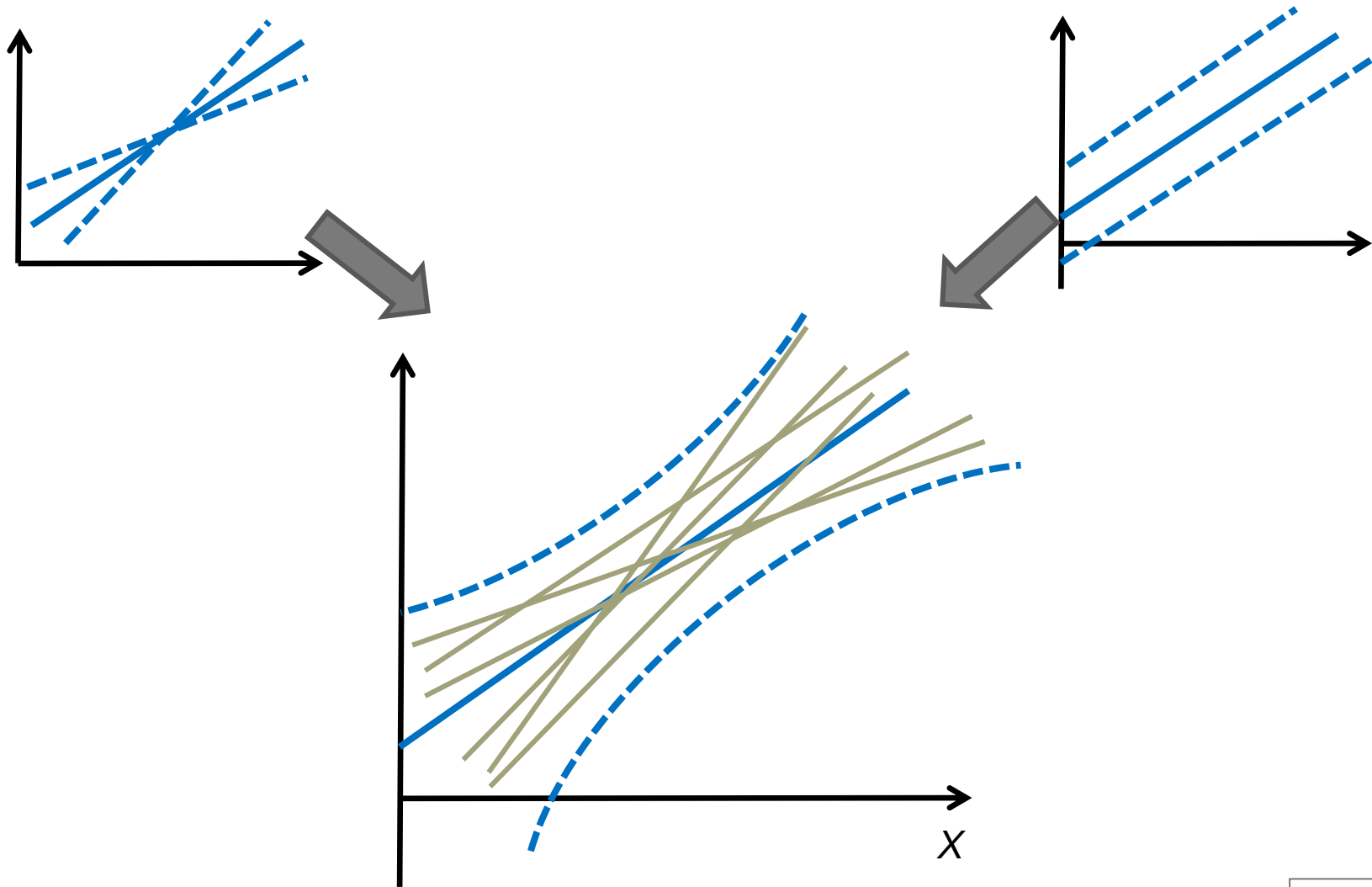


Vabaliige b

$$\Delta b = t_{\alpha/2}(\nu) se(b)$$



PARAMEETRITE USALDUSVAHEMIKE KOOSMÕJU



NÄIDE: AUTOTOOTJA BMW KULUFUNKTSIOON

Aluseks BMW näitajad 2004-2013: toodetud autode arv q ja kulud C (mln €) aastas.

Otsime kulufunktsiooni kujul $C = aq + b$

Regressioonanalüüs Excelis:

		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
b	Intercept	9304	1535	6,06	0,000302	5764	12844
a	Autod	0,0205	0,000993	20,7	3,16E-08	0,0182	0,0228

Parameetrid

Parameetrite
standardvead

Parameetrite
usalduspiirid

$$\hat{C} = 0,0205q + 9304$$

Piirkulu

Püsikulu

Piirkulu usaldusvahemik 0,0182 kuni 0,0228 mln €/tk.

Püsikulu usaldusvahemik 5764 kuni 12844 mln €.

ÜKSIKVÄÄRTUSE JA KESKVÄÄRTUSE PROGNOOSIMINE

Toidule tehtavate kulutuste y_T sõltuvus kogukuludest x

$$\hat{y}_T = 434,4 + 0,1297x$$

Kui palju kulutab pereliikme kohta aastas toidule pere, kellel kogukulud pereliikme kohta on 2000 € aastas?

$$\hat{y}_T(2000) = 434,4 + 0,1297 \cdot 2000 \approx 694 \quad \text{eurot aastas}$$

Prognoosi usaldusvahemik $\hat{y}(2000) = 694 \pm 324$ eurot aastas

ÜKSIKVÄÄRTUSE
PROGNOOS

Kui palju kulutavad **keskmiselt** pereliikme kohta aastas toidule **kõik** pered, kellel kogukulud pereliikme kohta on 2000 € aastas?

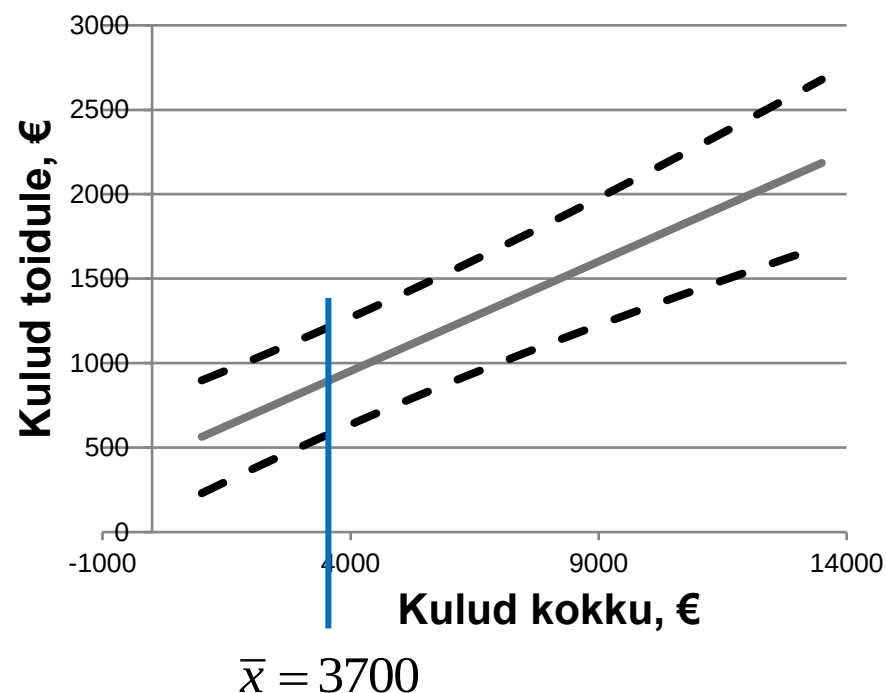
$$\hat{\mu}_T(2000) = 434,4 + 0,1297 \cdot 2000 \approx 694 \quad \text{eurot aastas}$$

Prognoosi usaldusvahemik $\hat{\mu}_T(2000) = 694 \pm 117$ eurot aastas

KESKVÄÄRTUSE
PROGNOOS

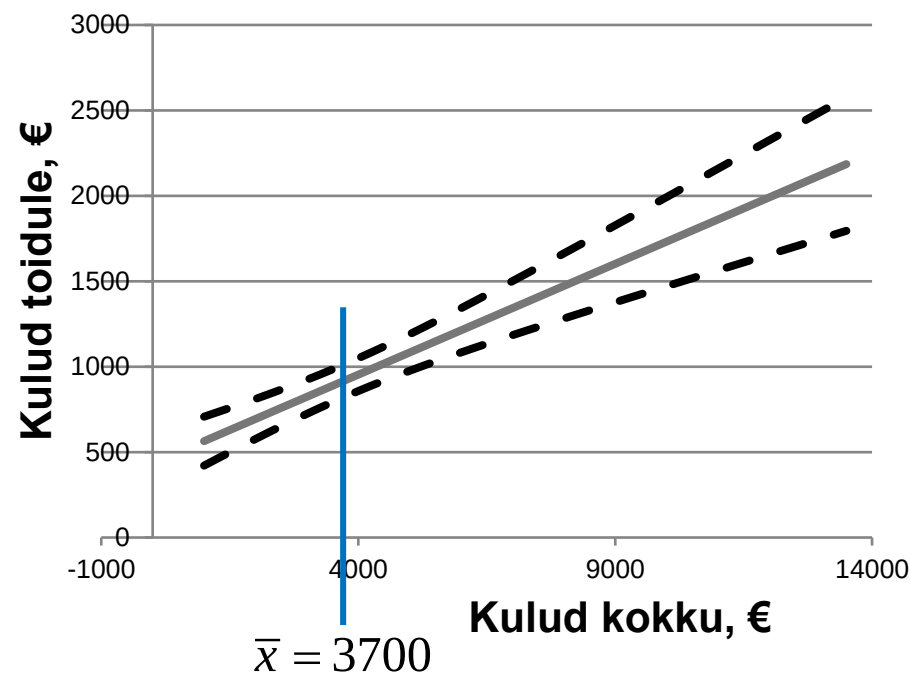
ÜKSIKVÄÄRTUSE JA KESKVÄÄRTUSE PROGNOOSIMINE

Üksikväärtuse prognoos ja
usalduspiirid



Valimi keskmine

Keskväärtuse prognoos ja
usalduspiirid



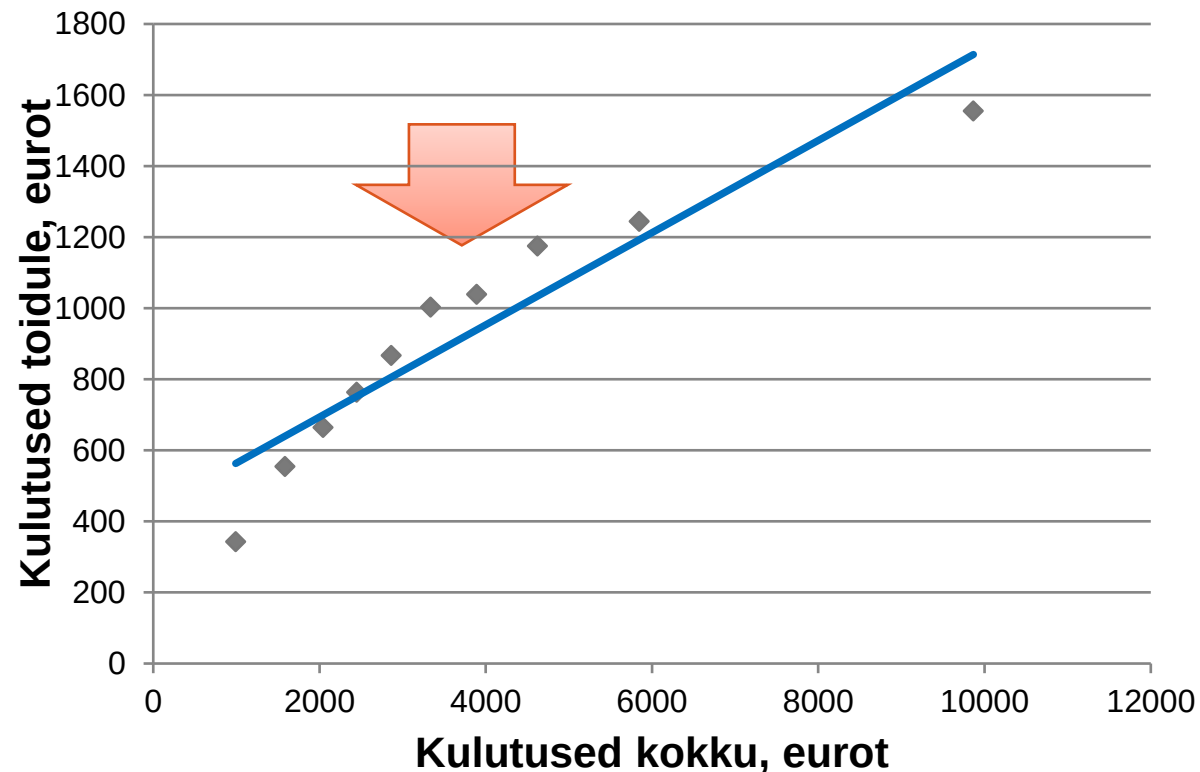
Valimi keskmine

Mida kaugemale valimi keskmisest, seda suurem on prognoosi viga.

NÄIDE: KULUD TOIDULE

$$\hat{y}_T = 434,4 + 0,1297x \quad R^2 = 0,881$$

Punktide kõrvalekaldumine sirgest pole juhuslik.

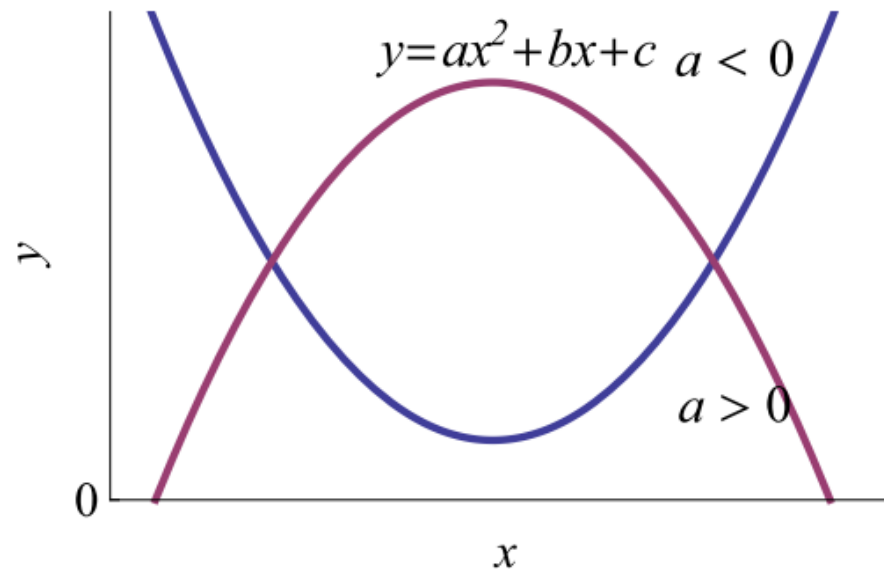


Kas sirgjoon on sobiv joon?

MUDELI KUJU

- Regressioonianalüüs võimaldab hinnata mudeli parameetrite arvväärtsi.
- Ei ütle, milline matemaatiline kuju peab mudelil olema.
- Lineaarne mudel $y = ax + b$ on üks võimalikest.
- Mudel võib olla ka mittelineaarne.

Näiteks $y = ax^2 + bx + c$



NÄIDE: KULUD TOIDULE

$$\hat{y}_T = 434,4 + 0,1297x$$

$$R^2 = 0,881$$

Ruutpolünoom

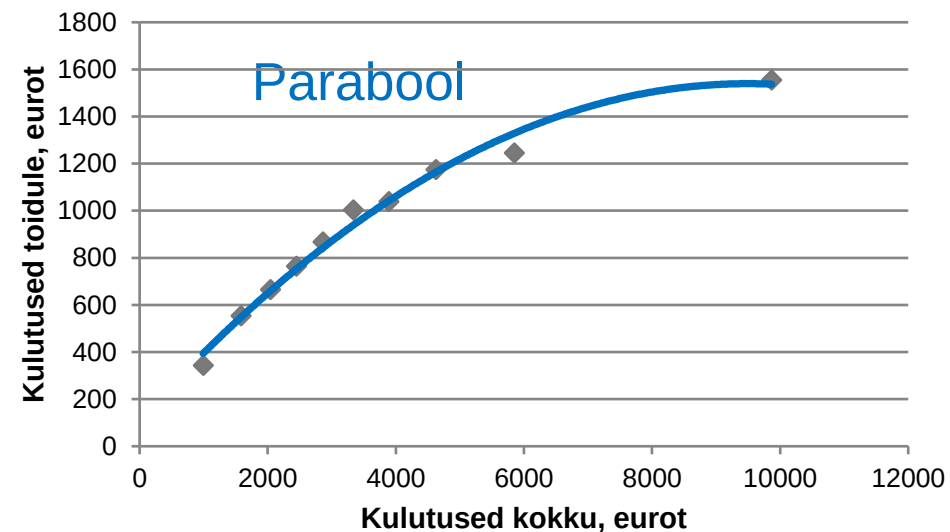
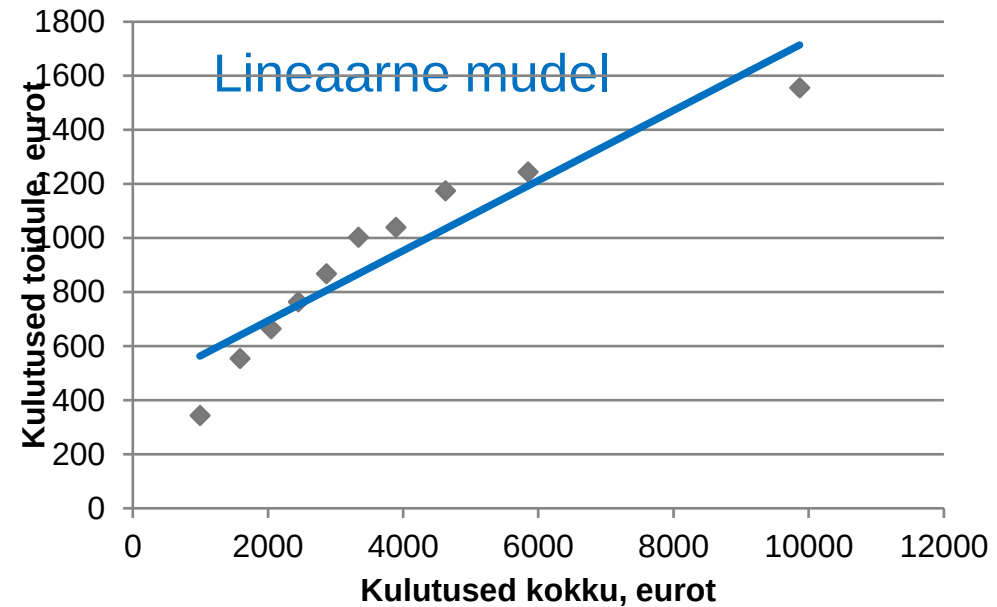
$$y = b + a_1x + a_2x^2$$

$$\hat{y}_T = 111,2 + 0,3x - 2 \cdot 10^{-5}x^2$$

$$R^2 = 0,987$$



On suurem



LINEAARNE VS MITTELINEAARNE

Lineaarse mudeli üldkuju $y = ax + b$

Demo: mudeli tõus

Muutumise kiirust näitab tuletis $y' = a$, on konstantne.

Kui x suureneb 1 võrra, siis y muutub a võrra iga x väärtuse korral.

Mittelineaarne mudel, näiteks parabool $y = ax^2 + bx + c$

Tuletis $y' = 2ax + b$ ei ole konstantne.

Kui x suureneb 1 võrra, siis y muutumine sõltub sellest, milline on x väärtus.

NÄIDE: KULUD TOIDULE LINEAARNE JA MITTELINEAARNE

Lineaarne mudel $\hat{y}_T = 434,4 + 0,1297x$

$$\Delta x = 1$$

$$x = 2000 \quad \Delta y = 0,1297$$

$$x = 6000 \quad \Delta y = 0,1297$$

Kui kogukulud (x) on 1 € võrra suuremad, siis kulud toidule (y) on suuremad 0,1297 € võrra iga x korral.

Muutus on konstantne.

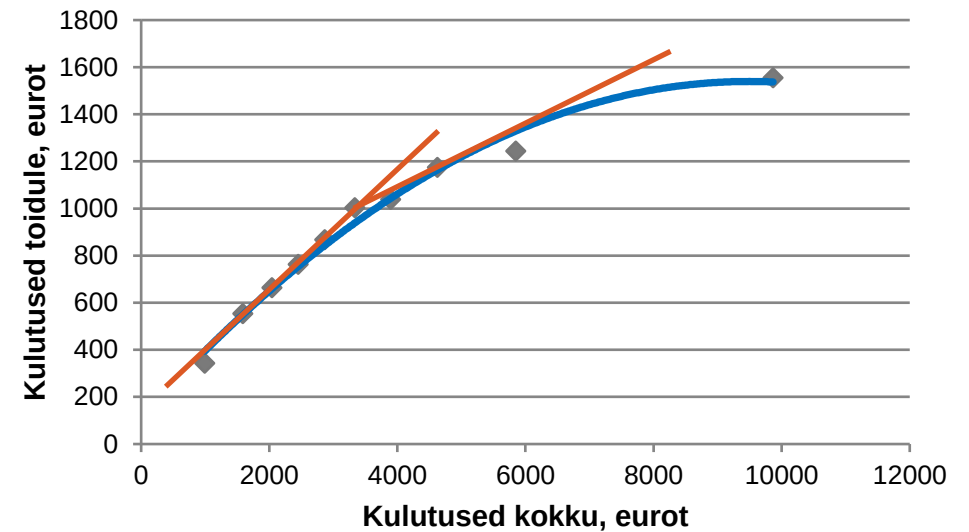
Mittelineaarne mudel $\hat{y}_T = 111,2 + 0,3x - 2 \cdot 10^{-5} x^2$

$$\text{Tuletis } \hat{y}_T' = 0,3 - 4 \cdot 10^{-5} x$$

$$\Delta x = 1$$

$$x = 2000 \quad \Delta y = 0,3 - 4 \cdot 10^{-5} \cdot 2000 \approx 0,22$$

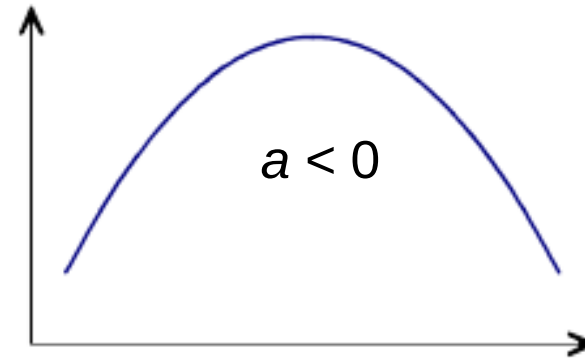
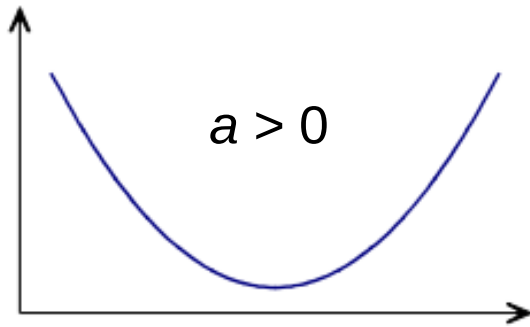
$$x = 6000 \quad \Delta y = 0,3 - 4 \cdot 10^{-5} \cdot 6000 \approx 0,06$$



Kasv aeglustub.

KÕVERUSE INTERPRETEERIMINE

$$y = ax^2 + bx + c$$

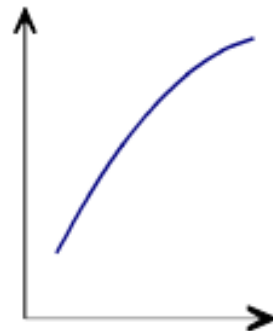


Kasvamine

Kahanemine



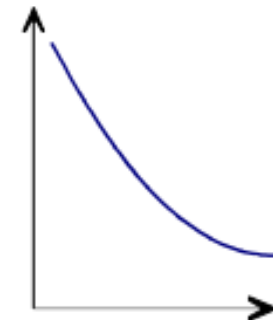
kiirenev
 $a > 0$



aeglustuv
 $a < 0$



kiirenev
 $a < 0$



aeglustuv
 $a > 0$

???

Mikroökonomikast on kulufunktsiooni kohta teada, et kui tootmismaht muutub väga suurtes piirides, ei pruugi kulufunktsioon olla lineaarne.

Analüüsi 145 USA elektrijaama kulusid ja elektri tootmise mahtu. Saadi, et kulufunktsioon on

$$\hat{C} = 1,88 + 0,0037q + 2,4 \cdot 10^{-7} q^2$$

kus C on elektrijaama kogukulud (mln \$) ja q tootmismaht (mld kWh).

Kas tootmismahu suurenedes piirkulu (lisakulu lisaühiku tootmisel)

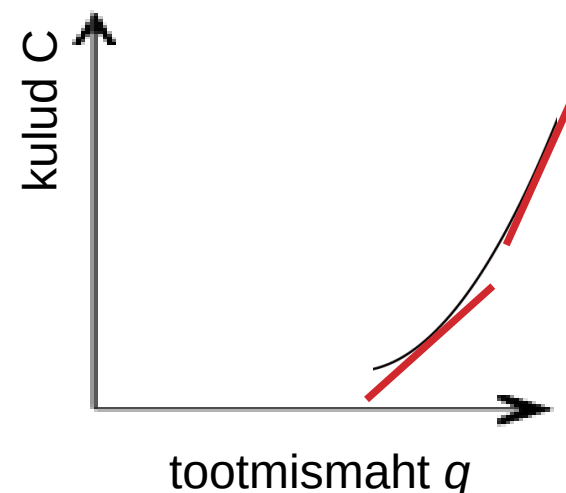
- jääb samaks;
- väheneb;
- suureneb?

V

1. Ruutliikme kordaja > 0 , järeltõusav parabool.

2. Kuna tootmismahu kasvades kulud saavad vaid kasvada, järeltõusav kasv.

3. Piirkulu on puutuja tõus. Suurema tootmismahu korral on tõus suurem.



NÄIDE: PARABOOLI EKSTREEMUMKOHT

Andmed Eesti Statistikaameti Sotsiaaluuringust 2013.

Töötavad isikud vanuses 20 kuni 70 aastat, kokku 6143 isikut.

V vanus aastates

P rahaline sissetulek palgatööst aastas, eurot.

Palga sõltuvus vanusest: $P_i = -26,9 + 409V_i - 5,05V_i^2 + u_i$

Millise vanuse korral on palk kõige suurem?

Parabooli maksimumkoha leidmine: leida tuletis ja panna see võrduma nulliga.

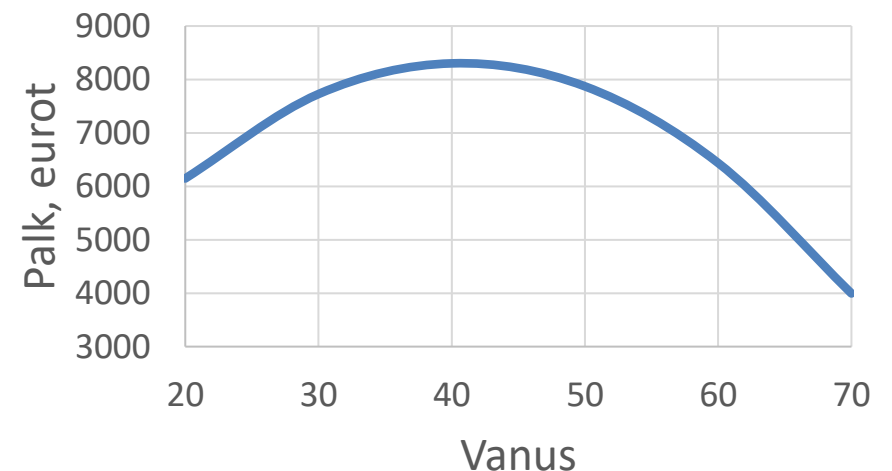
$$P'_i = 409 - 2 \cdot 5,05V$$

$$P'_i = 0$$

$$409 - 2 \cdot 5,05V = 0$$

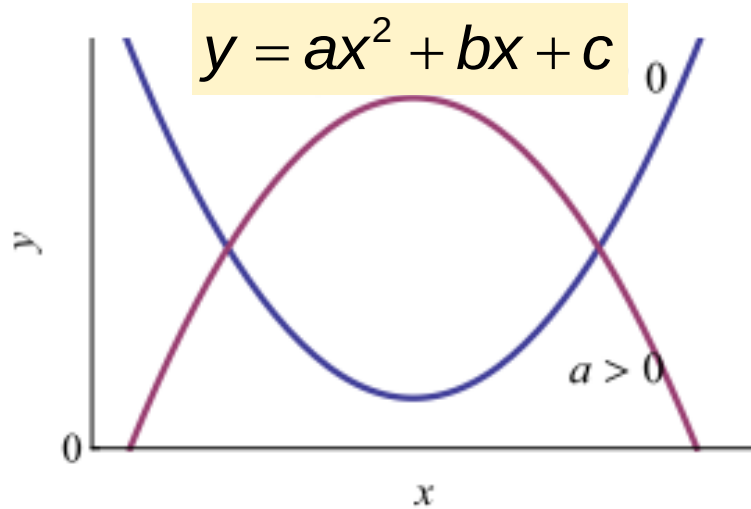
$$V \approx 40$$

Palk tõuseb kuni vanuseni 40 ja siis hakkab langema.

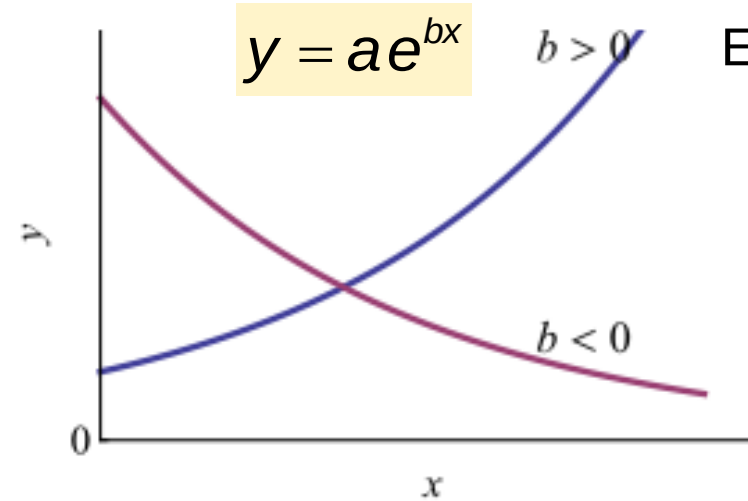


TÜÜPILISEMAD MITTELINEAARSED MUDELID

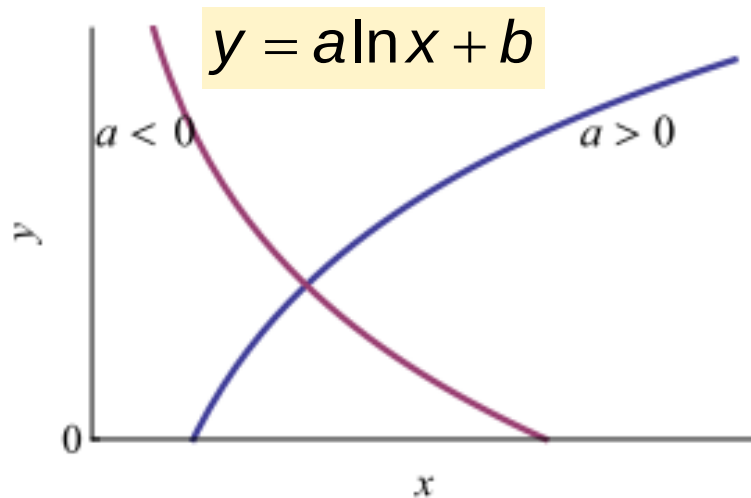
Parabool



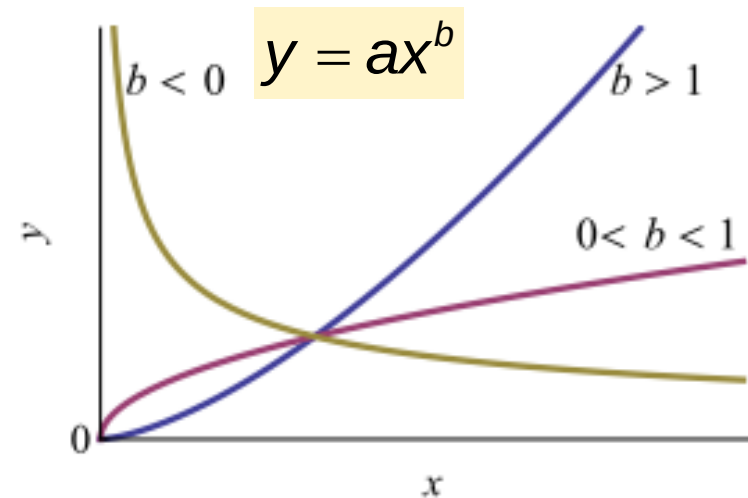
Eksponentsiaalne



Logaritmiline



Astmefunktsioon



MUDELI KUJU VALIK

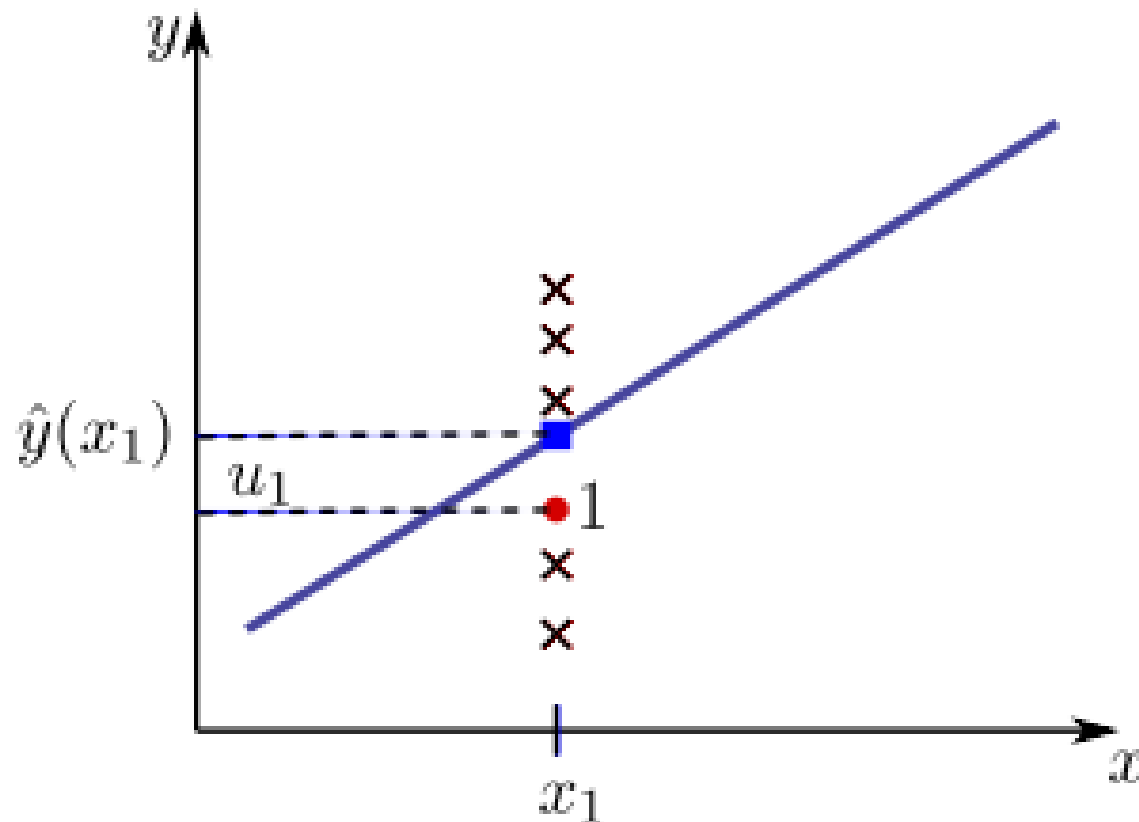
- Teooriast **on teada**, millise kujuga seos uuritavate suuruste vahel eksisteerib, on teada mudeli üldkuju.
 - On vaja leida vaid konkreetsele andmestikule vastavad parameetrid, st mudeli konkreetne kuju.
- Seose kuju **pole teada**.
 - Proovitakse erineva kujuga mudeleid ning determinatsioonikordaja järgi otsustatakse, milline neist kirjeldab seda seost kõige paremini (determinatsioonikordaja on kõige suurem).

Järelikult võimaldab regressioonanalüüsi läbiviimine täiendada ka majandusteooriat, pakkudes välja empiirilise mudeli uuritava seose kirjeldamiseks.

REGRESSIOONI JÄÄGID

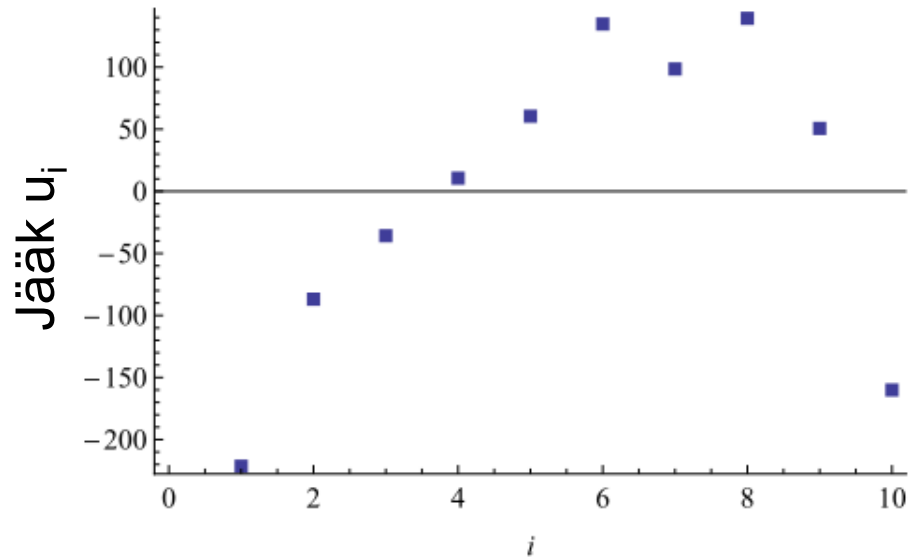
Regressiooni jääk on valimisse kuuluva objekti tegeliku väärtuse ja mudelväärtuse vahe.

$$u_i = y_i - \hat{y}_i$$

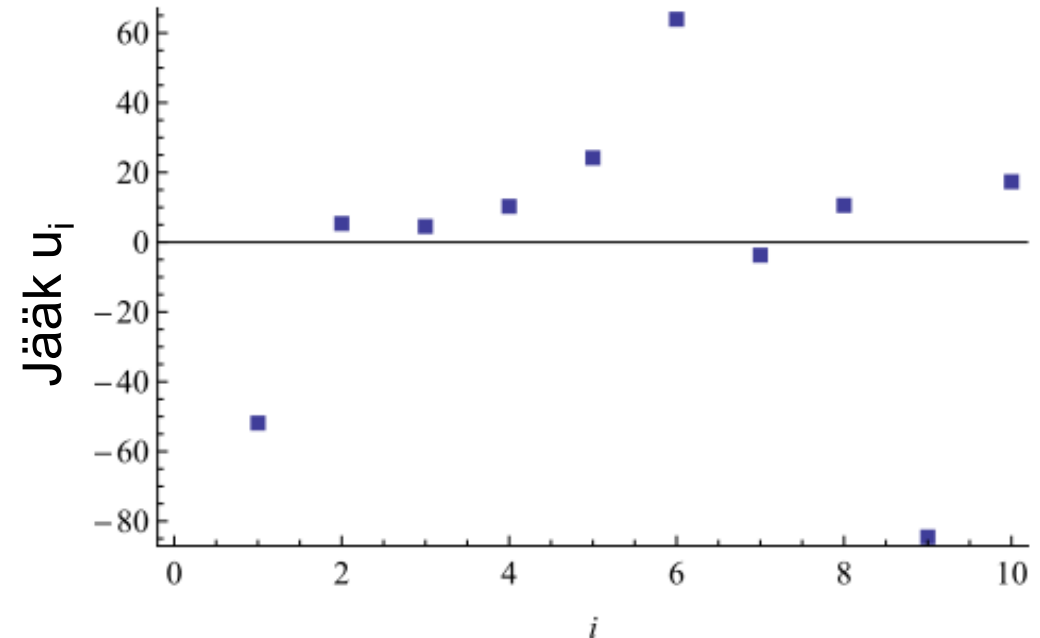


NÄIDE: REGRESSIOONI JÄÄGID JA KULUD TOIDULE

Lineaarse mudeli jäägid



Paraboolse mudeli jäägid



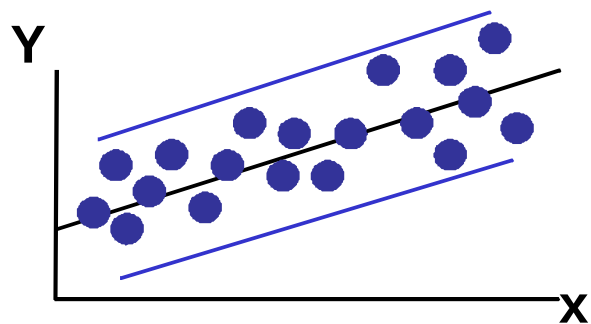
Jäägid paiknevad juhuslikumalt.

JÄÄKIDE ANALÜÜS

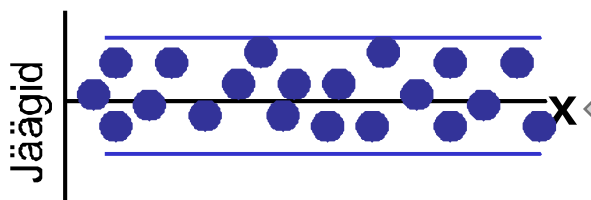
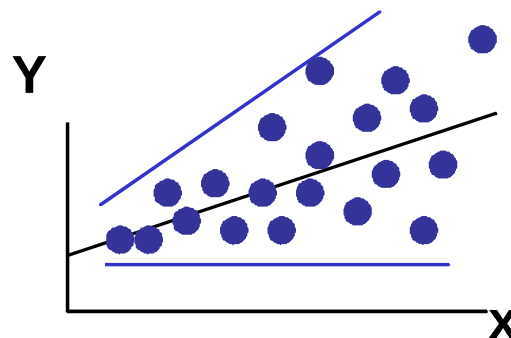
Nõuded regressioonmudeli jääkidele:

- 1) jäägid on normaaljaotusega, mille keskväärtus on 0;
- 2) jääkide hajuvus on konstantne;
- 3) jäägid on sõltumatud argumenttunnusest.

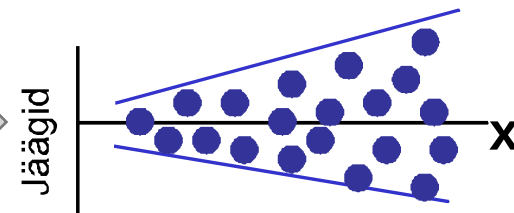
PEAB OLEMA



EI TOHI OLLA



Jääkide
diagrammid



STANDARDISEERITUD JÄÄGID JA ERINDID

Standardiseeritud jääk saadakse jäägi jagamisel jääkide standardhälbega:

$$u_i^{std} = \frac{u_i}{s_{n-1}}, \quad s_{n-1} = \sqrt{\frac{1}{n-1} \sum u_i^2}$$

Võimaldavad välja tuua erindeid.

Esinemise tõenäosus.

$$|u_i^{std}| > 2 \quad \text{ebatüüpiline vaatlus} \quad p < 0,05$$

$$|u_i^{std}| > 3 \quad \text{erind} \quad p < 0,01$$

NÄIDE: SKP JA LAUATELEFONIDE ARV

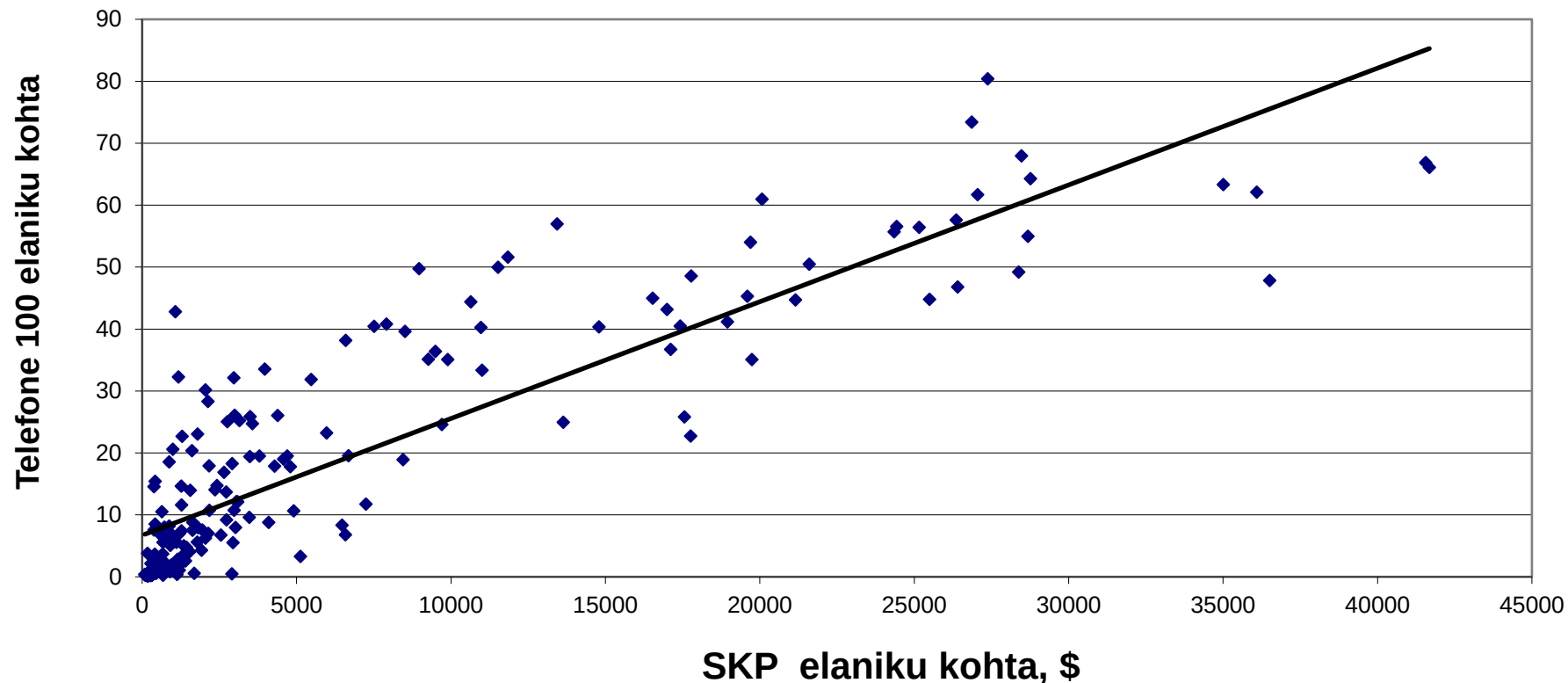
Andmed 138 riigi kohta 1998. aastal:

x SKP elaniku kohta;

y lauatelefonide arv 100 elaniku kohta.

$$\hat{y} = 0,0017x + 10,2$$

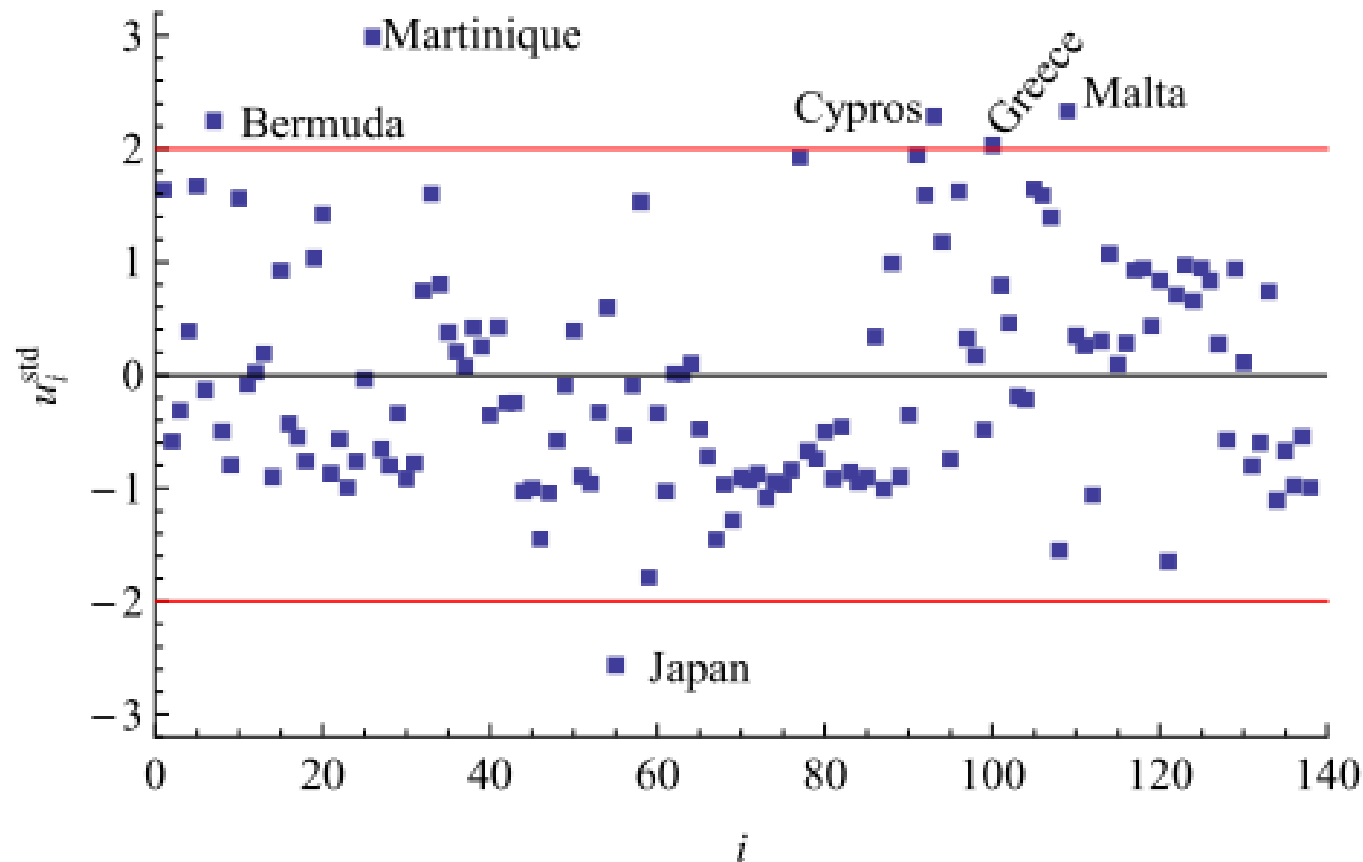
$$R^2 = 0,75$$



NÄIDE: SKP JA LAUATELEFONIDE ARV

Punased jooned vastavad standardiseeritud jäägile -2 ja 2.

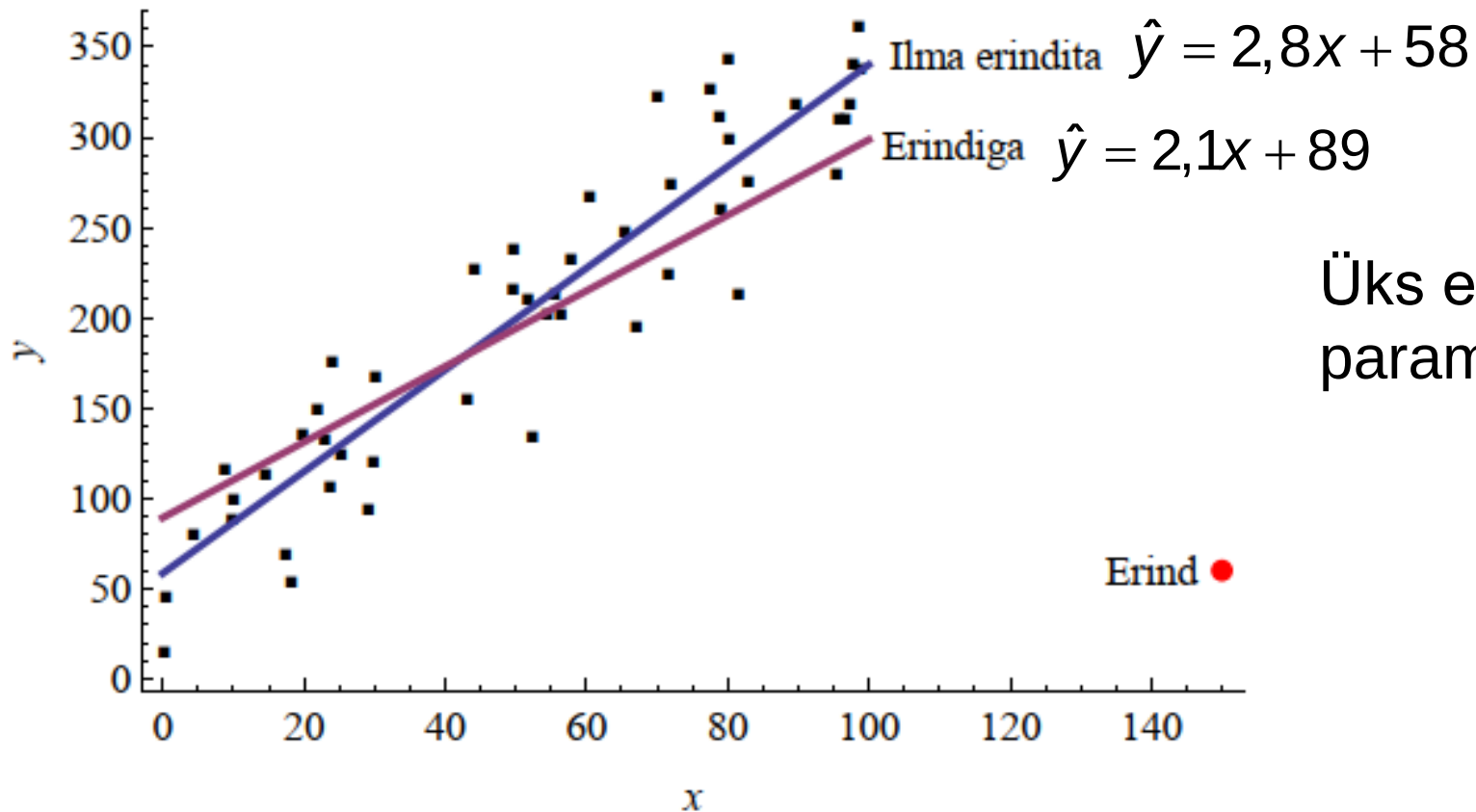
Jääkide
diagramm



Ebatüüpilisi riike on 6.

ERINDI MÕJU

Demo: erindi mõju



Üks erind võib oluliselt mõjutada parameetrite hinnanguid.

Vähimruutude meetodil leitud mudeli parameetrid on tundlikud erindite suhtes.

REGRESSIOONMUDELI KASUTAMINE

- Mudel kirjeldab suuruste vahelist seost:
 - mis suunas üks suurus teist mõjutab;
 - kui palju mõjutab;
 - kas mõju on lineaarne või mittelineaarne.
- Mudel võimaldab prognoosimist.
- Mudel võimaldab välja tuua erindeid.