



Valsts administrācijas
skola

Datu analīze – 2.līmenis

Edgars Brēķis, 4.nodarbība

Regresija:

Lineārā vienfaktora regresija

Lineārā daudzfaktoru regresija

Lineārās regresijas diagnostika

Mācības finansētas no Eiropas Sociālā fonda projektiem

Nr.3.4.2.0/15/I/001 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide labāka regulējuma izstrādē mazo un vidējo komersantu atbalsta jomā" un

Nr.3.4.2.0/15/I/002 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide korupcijas novēršanas un ēnu ekonomikas mazināšanas jomā"

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA

Eiropas Sociālais
fonds

I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē



Valsts administrācijas
skola

Mācību plānojums, 4.nodarbība

9.00-11.00	Ieskats regresiju analīzē. Mazāko kvadrātu metode. Regresijas vienādojuma kvalitātes novērtējums.
11.00-11.15	Kafijas pauze
11.15-12.15	Determinācijas koeficients, koriģētais determinācijas koeficients. Situācijas analīze. Izlases apjoms. Netipisko vērtību analīze. Neatkarīgo mainīgo analīze. Modeļa specifikācija.
12.15-13.00	Pārtraukums
13.00-14.30	Regresijas analīzes klasiskie pieņēmumi. Galveno ekonometrisku problēmu konstatēšana un novēršana. Situācijas analīze.
14.30-14.45	Kafijas pauze
14.45-16.15	Regresijas projekts ar SPSS



Valsts administrācijas
skola

Ieskats regresiju analīzē

Mērķis, pamatjēdzieni



Regresijas analīzes pārskats

- Kas ir ekonometrija?
 - Ekonometrija literāri nozīmē «ekonomikas mērījumi»
 - Tā ir aktuālo ekonomikas un biznesa fenomenu kvantitatīva novērtēšana un analīze, kas sevī ietver:
 - a) ekonomikas teoriju,
 - b) statistiku,
 - c) matemātiku un
 - d) novērojumu/datu vākšanu.
 - Ekonometrijas trīs pamatpielietojumi:
 - 1) aprakstīt ekonomikas realitāti,
 - 2) pārbaudīt ekonomikas teorijas hipotēzes,
 - 3) prognozēt ekonomikas aktivitāti.



Piemērs: ekonomikas teorijas saistība ar ekonometriju

- Aplūkosim teorētisku sakarību starp vietējo un ārvalstu cenu līmeņiem, ienākumu līmeni un vietējo preču un pakalpojumu pieprasījumu:
$$Q = f(P, P_f, Y_d)$$
- Pirms sakarības novērtēšanas, teorijā jāatrod atbilde:
 - kādas zīmes sagaidām novērtējamiem koeficientiem?
- Ekonometrijas metodes, ļauj teorētisko sakarību izteikt kā vienādojumu:
 - $Q = 27.7 - 0.11P + 0.03P_s + 0.23Y_d$
- Jautājumi:
 - vai novērtēto koeficientu zīmes atbilst teorētiski sagaidāmām?
 - vai modelis (vienādojums) kopumā ir statistiski nozīmīgs (izskaidro pieprasījuma dažādību)?
 - vai novērtētie koeficienti ir statistiski nozīmīgi?
 - kāda ir neatkarīgo mainīgo izmaiņu ietekme uz pieprasījumu?



Kas ir regresijas analīze?

- Regresijas analīze ir statistiskas metode, kas «izskaidro», viena mainīgā izmaiņas, **atkarīgā mainīgā**, kā funkciju no citu mainīgo izmaiņām (neatkarīgiem, vai izskaidrojošiem mainīgajiem), ar viena vienādojuma palīdzību.

- Piemērā

$$Q = f(P, P_f, Y_d)$$

Q – atkarīgais mainīgais

P, P_f , Y_d – neatkarīgie, jeb izskaidrojošie mainīgie

NB Statistiski nozīmīgs regresijas vienādojums vēl nenozīmē cēloņsakarību, to ir jāapstiprina arī ekonomikas teorijai un ... loģikai.



Lineāri viena-vienādojuma modeļi

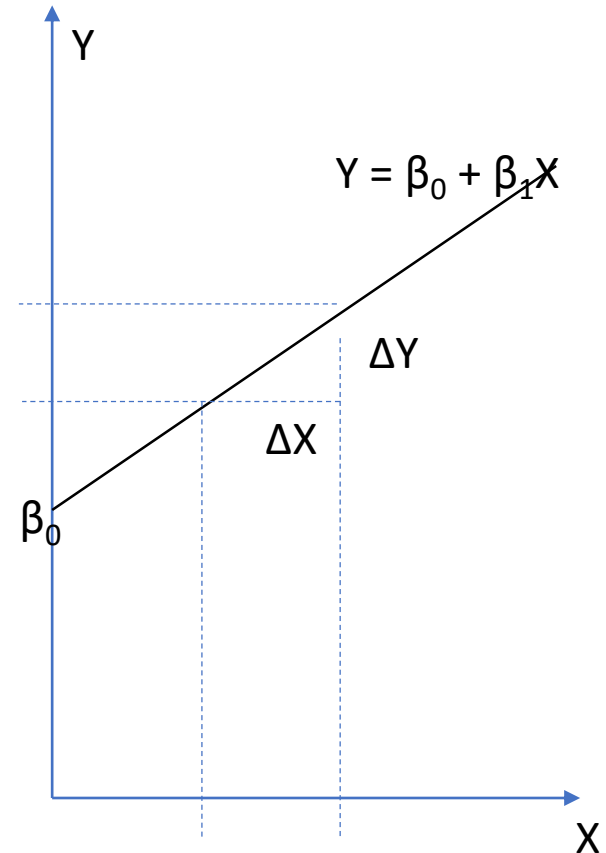
- Vienkāršākais piemērs ir vienfaktora regresija:
- $Y = \beta_0 + \beta_1 X$
- β -as sauc par «**koeficientiem**»
 - β_0 ir «**konstante**», jeb «**brīvais koeficients**»
 - β_1 ir «**virziena koeficients**»: par cik vienībām izmainīsies Y , ja X tiks palielināts par vienu vienību.

$$Y = \beta_0 + \beta_1 X \quad (1) \text{ dotais vienādojums}$$

$$Y + \Delta Y = \beta_0 + \beta_1 (X + \Delta X) \quad (2) \text{ izmainot } X, \text{ izmainās arī } Y$$

$$\Delta Y = \beta_1 \Delta X \quad \text{no (2) atņem (1)}$$

$$\beta_1 = \Delta Y / \Delta X \quad \text{taisnes virziena koeficients}$$





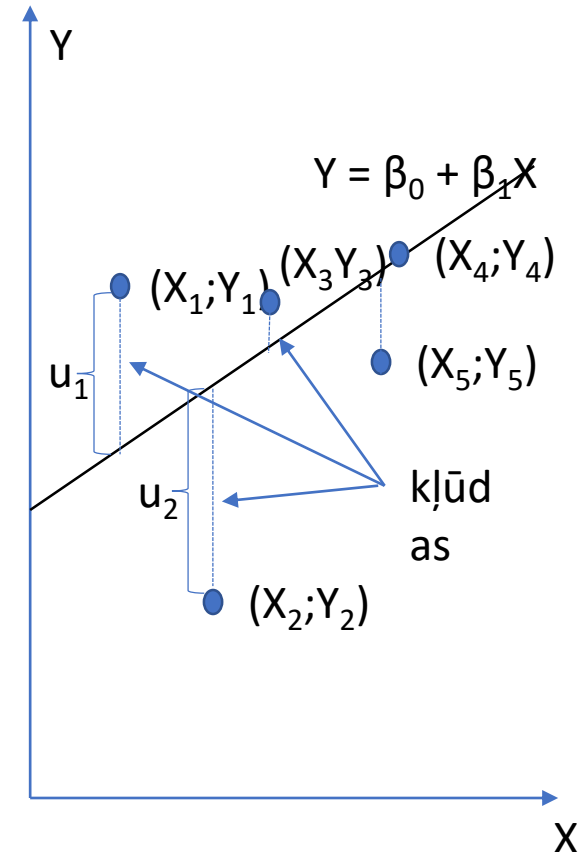
Lineāri viena-vienādojuma modeļi

- Par to, ka $Y = \beta_0 + \beta_1 X$ ir lineārs modelis, šaubu nav! Vai lineārs ir arī modelis $Y = \beta_0 + \beta_1 X^2$?
 - To, vai modelis ir lineārs, secina pēc nezināmo sakarības ar izskaidrojamo mainīgo!
 - Kas modelī ir nezināmie? Y un X ir pētāmie dati, tātad zināmi lielumi. Nezināmie, tas ko jānovērtē, ir koeficienti!
- Modelis ir lineārs, ja atkarīgo mainīgo Y izskaidro modeļa koeficientu lineāra kombinācija – koeficientu un mainīgo reizinājuma summa vai starpība.
 - $Y = \beta_0 + \beta_1 X^2$ ir lineārs (ja apzīmē $X^2 = Z$, tad $Y = \beta_0 + \beta_1 Z$)
 - $Y = \beta_0 \times \beta_1 X$ nav lineārs
 - $Y = \beta_0 X^{\beta_1}$ nav lineārs



Lineāri viena-vienādojuma modeļi

- Vai lineārais modelis $Y = \beta_0 + \beta_1 X$ apraksta visas Y dažādās vērtības?
Atbilde ir, nē, jo:
 - iespējams ir vēl citi izskaidrojošie mainīgie, ne tikai X , kas ietkmē Y vērtības;
 - mērījuma kļūdas;
 - nepareizi izvēlēts funkcijas veids;
 - Y ir gadījuma rakstura, kuru neietekmē X .
- Lai atspoguļotu Y variācijas, kuras neapraksta mainīgais X , vienādojumā tiek pievienota «stohastiska (gadījuma rakstura) kļūdas» komponente u :
$$Y = \beta_0 + \beta_1 X + u$$





Lineāri viena-vienādojuma modeļi

- Vienādojumu $Y = \beta_0 + \beta_1 X + u$ iedala divās komponentēs:
 - deterministiskā (noteicošā) komponente ($\beta_0 + \beta_1 X$) un
 - stohastiskā (gadījuma rakstura) komponente (u)
- Kāpēc «deterministiska»?
 - tā parāda kāda ir sagaidāmā Y vērtība pie dotās X vērtības, $E(Y|X)$
 - t.i., ja pie izvēlētās X vērtības ir novērojamas vairākas Y vērtības, tad kāda vidējā Y vērtība atbilst konkrētai X vērtībai



Par apzīmējumiem - indeksi

- Regresijas vienādojuma mainīgiem tiek pievienoti indeksi, tā uzsverot, ka tiek izmantoti izlases novērojumi.
 - i – ja tiek izmantoti individuālo novērojumu dati (šķērsgriezuma dati)
 - t – ja tiek izmantotas laika rindas
 - it – ja izmanto abus, individuālos un laika rindu datus, t.i., paneļdatus.
- $Y_i = \beta_0 + \beta_1 X_i + u_i$ ($i = 1, 2, \dots, 750$) individuālie novērojumi, izlases apjoms 750
- $Y_t = \beta_0 + \beta_1 X_t + u_t$ ($t = 2012.01, \dots, 2017.08$) laika rinda, mēnešu dati periodā no 2012.g. janvāra līdz 2017.g. augustam, novērojumu skaits 80
- $Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$ ($i = 1, 2, \dots, 750$ un $t = 2012.01, \dots, 2017.08$) paneļdati, 750 indivīdiem novērojumi apkopoti par 80 mēnešu periodu



Par apzīmējumiem – novērtētais regresijas vienādojums

- Līdz šim aplūkotais regresijas vienādojums $Y_i = \beta_0 + \beta_1 X_i$ ir «patiesss», bet vispārīgs – teorētisks regresijas vienādojums
- Tā vietā, lai iegūtu «patiesu» ģenerālkopas regresiju, vienādojums tiek novērtēts izmantojot tikai kādu ģenerālkopas izlasi.
- Empīriski novērtēto regresijas vienādojumu pieraksta

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

- kur «jumiņi» norāda, ka parametri ir novērtēti (aprēķināti), izmantojot doto izlasi.
 - Ja novērtētā Y vērtība $\hat{Y}_i$ ir tuvāk novērotai Y vērtībai Y_i , jo labāk vienādojums apraksta datus, t.i.,
 - mazākas «klūdas» $u_i = Y_i - \hat{Y}_i$, labāk vienādojums apraksta datus.



Valsts administrācijas
skola

Mazāko kvadrātu metode



Vienfaktora modeļa novērtēšana ar mazāko kvadrātu metodi (MKM)

- Mērķis ir novērtēt regresijas vienādojuma $Y_i = \beta_0 + \beta_1 X_i + u_i$ parametrus, β_0 un β_1 , lai iegūtu vienādību $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$
- Plašāk pielietotā metode ir Mazāko kvadrātu metode (MKM)
- MKM minimizē $\sum_{i=1}^N u_i^2$, t.i., minimizē kļūdu kvadrātu summu
 - minimizē vertikālo attālumu kvadrātu summu no novērojuma punkta līdz punktam uz taisnes $\sum_{i=1}^N (Y_i - \hat{Y}_i)^2$
- Kāpēc izmantot MKM:
 - pieejama visās standarta statistikas paketēs, arī EXCEL
 - intuitīvi viegli saprotama - mērķis ir atrast vienādojumu, kas būtu pēc iespējas tuvāks datiem
 - MKM metode ļauj aprēķināt statistikas, kas noder modeļa atbilstības novērtēšanai



Kā darbojas MKM vienfaktora regresijas modelī

- Matemātiski var pierādīt
- parametri, kas nodrošina, ka kļūdu kvadrātu summa ir vismazākā, aprēķināmi ar sekojošu vienādojumu palīdzību

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N [(X_i - \bar{X}) (Y_i - \bar{Y})]}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$



Daudzfaktoru regresijas modeļa novērtēšana ar MKM

- «Reālajā pasaulē» ar vienu izskaidrojošo mainīgo var nepietikt, lai aprakstītu Y variācijas.
- Vispārīgais daudzfaktoru regresijas modelis ar K neatkarīgiem mainīgajiem ir:
- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + u_i \quad (i = 1, 2, \dots, N)$
- Lielākā atšķirība ir tikai virziena koeficientu interpretācijā:
 - Tagad virziena koeficients norāda uz atkarīgā mainīgā Y izmaiņām, ja par vienu vienību palielina kāda neatkarīgā mainīgā vērtību, pārējiem mainīgajiem paliekot fiksētiem
 - Kļūdu kvadrātu minimizēšanas principi nemainās
 - Nenovērotie (un tāpēc modelī neiekļautie) mainīgie netiek kontrolēti, un tāpēc netiek fiksēti



Kopējā, izskaidrotā un neizskaidrotā kvadrātu summa

Kopējā kvadrātu summa

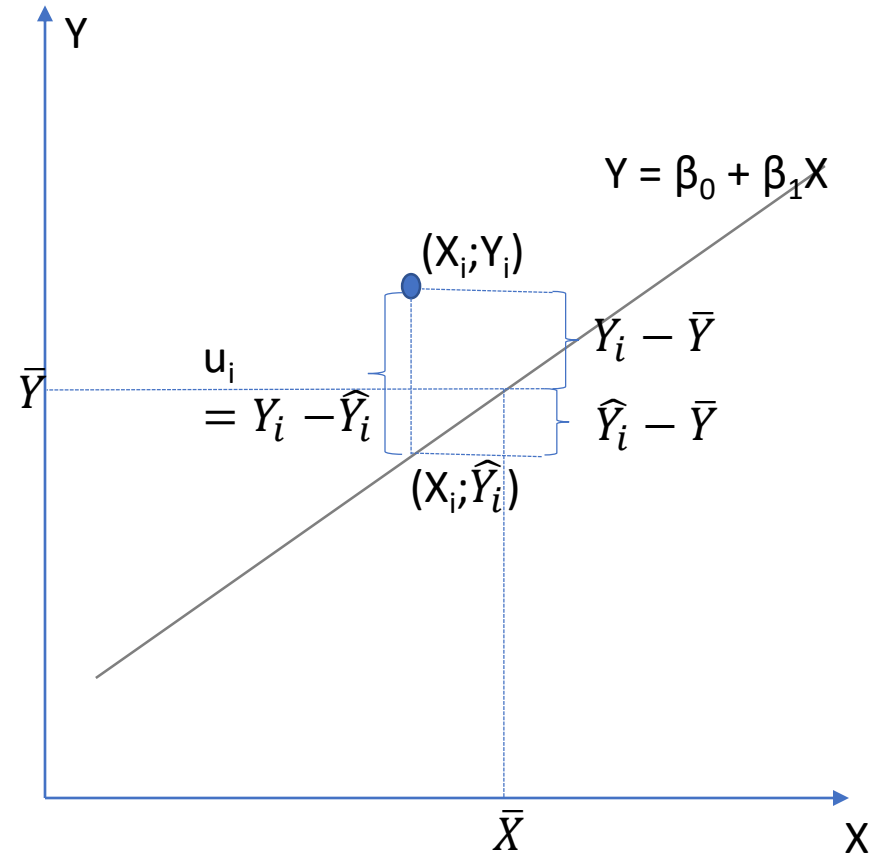
$$TSS = \sum_{i=1}^N (Y_i - \bar{Y})^2 =$$

$$= \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2 +$$

izskaidrotā kvadrātu
summa (ESS)

$$+ \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

neizskaidrotā kvadrātu
summa (RSS)





Regresijas vienādojuma kvalitātes novērtējums

- Jāpārbauda sekojošo:
 - ☐ Vai vienādojums ir **teorētiski pamatots**?
 - ☐ Cik labi novērtētais vienādojums **atbilst datiem**?
 - ☐ Vai datu **izlase** ir pietiekami **liela un «tīra»**?
 - ☐ Vai dotam vienādojumam **MKM** ir **piemērotākā metode**?
 - ☐ Kā novērtētie parametri atbilst **gaidām**, kuras izveidojās pirms uzsāka datu vākšanu?
 - ☐ Vai **visi** acīmredzami **svarīgie mainīgie** ir **iekļauti** vienādojumā?
 - ☐ Vai ir izmantota teorētiski visloģiskākā **funkcijas forma**?
 - ☐ Vai regresija ir **brīva no** galvenajām **ekonometrijas problēmām**?

Pārbaudāmie punkti, ir tēmas, «Regresijas», satura rādītājs!



Valsts administrācijas
skola

Vai vienādojums ir **teorētiski pamatots**?

Cik labi novērtētais vienādojums atbilst datiem?

Vai datu **izlase** ir pietiekami **liela un «tīra»**?

Vai dotam vienādojumam **MKM** ir **piemērotākā metode**?

Kā novērtētie parametri atbilst **gaidām**, kuras izveidojās pirms uzsāka datu vākšanu?

Vai **visi** acīmredzami **svarīgie mainīgie** ir **iekļauti** vienādojumā?

Vai ir izmantota teorētiski visloģiskākā **funkcijas forma**?

Vai regresija ir **brīva no** galvenajām **ekonometrijas problēmām**?

Determinācijas koeficients



Determinācijas koeficients

- Determinācijas koeficients ir izskaidroto kvadrātu summas un kopējās kvadrātu summas attiecība:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^N u_i^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

- Mazākās kvadrātu metodes pamatprincips ir minimizēt kļūdu kvadrātu summu, tāpēc lielāka determinācijas koeficienta vērtība atbilst labākam modelim.

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

- Determinācijas koeficients raksturo, kādu daļu no kopējām Y variācijām izskaidro novērtētais regresijas vienādojums.



Daudzfaktoru regresijas modeļi: koriģētais determinācijas koeficients

- Modelim pievienojot papildus izskaidrojošos mainīgo (X) R^2 palielinās, jo tiek izskaidrota vēl kāda daļa no kvadrātu summām.
- Vai tāpēc modelis kļūst labāks?
- Daudzfaktoru regresijas modeļiem, lai novērtētu novērtētā modeļa atbilstību datiem, R^2 vietā, tiek izmantots koriģētais determinācijas koeficients $\bar{R}^2$:

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^N u_i^2 / (N - K - 1)}{\sum_{i=1}^N (Y_i - \bar{Y})^2 / (N - 1)}$$

kur N – izlases apjoms, K- neatkarīgo

mainīgo skaits

- Koriģētais determinācijas koeficients dod iespēju salīdzināt modeļus ar dažādu izskaidrojošo mainīgo skaitu.



Valsts administrācijas
skola

Situācijas analīze

Datu failā «dati_ipasums» apkopotas 321 nekustamā īpašuma cenas un tos raksturojošie rādītāji: īpašuma vecums, ēkas un zemes platības, istabu skaits, vannasistabu skaits, palīgēku skaits, attālums līdz Rīgai

Izmantojiet dotos datus, statistikas programmu SPSS, atbildiet uz sekojošiem jautājumiem

- Novērtējiet regresijas modeli, kurā īpašuma cenas izskaidrotu ēkas platība un istabu skaits.
- Uzrakstīt novērtēto regresijas vienādojumu.
- Vai novērtēto koeficientu zīmes atbilst sagaidāmajām?
- Vai modelis kopumā ir statistiski nozīmīgs?
- Vai novērtētie koeficienti ir statistiski nozīmīgi?
- Par cik mainīsies īpašuma cena, ja istabu skaits pieaug, bet platība nemainās?
- Ja ēkas plātība ir 180km^2 un tajā ir 4 istabas, kāda ir prognozējamā cena?
- Kādu daļu no cenu variācijām izskaidro ēkas platība un istabu skaits?
- Kuru mainīgo pievienošana uzlabotu cenu variācijas izskaidrospēju?



Valsts administrācijas
skola

Vai vienādojums ir **teorētiski pamatots**?

Cik labi novērtētais vienādojums **atbilst datiem**?

Vai datu **izlase** ir pietiekami **liela un «tīra»**?

Vai dotam vienādojumam **MKM** ir **piemērotākā metode**?

Kā novērtētie parametri atbilst **gaidām**, kuras izveidojās pirms uzsāka datu vākšanu?

Vai **visi** acīmredzami **svarīgie mainīgie ir iekļauti** vienādojumā?

Vai ir izmantota teorētiski visloģiskākā **funkcijas forma**?

Vai regresija ir **brīva no** galvenajām **ekonometrijas problēmām**?



Valsts administrācijas
skola

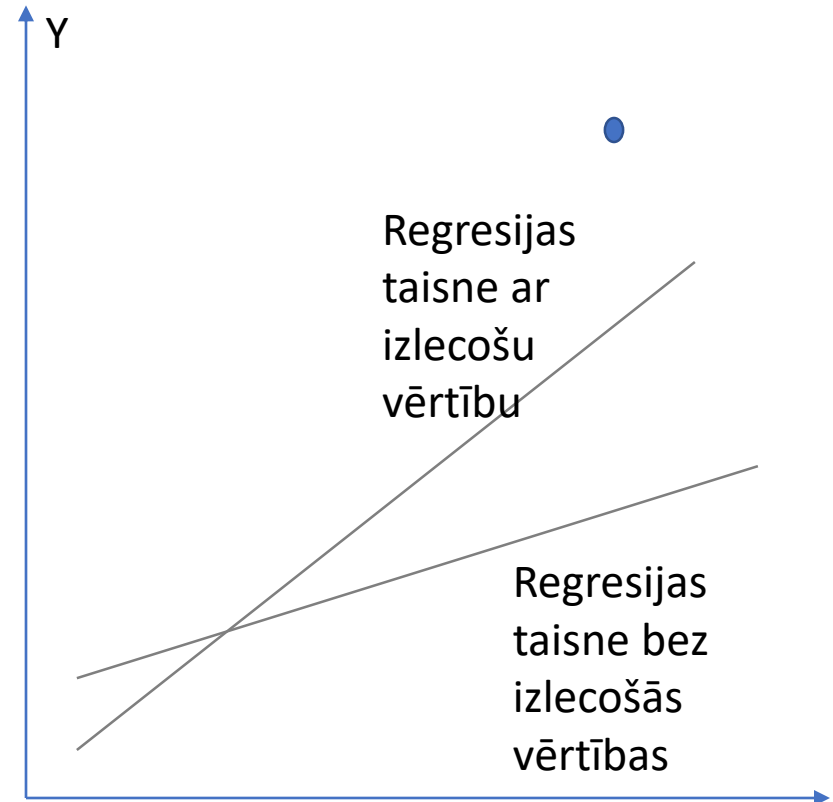
Izlsases apjoms

- Izlsases veidošanā jāievēro **statistikas pamatprincipi**, jā rūpējas, lai tā pietiekami labi reprezentētu ģenerālkopu.
- No modelēšanas viedokļa, izlsases apjoms ieteicams (!) 5-10 reizes lielāks, nekā neatkarīgo mainīgo skaits.



Netipisko vērtību analīze

- Izlecošie dati, netipiskas X vai Y vērtības var būtiski ietekmēt regresijas rezultātus.
- Jāveic izlecošo datu analīze! Vai tā ir pārrakstīšanās kļūda? Vai punkts pieder izlasei? Kāpēc radusies netipiska vērtība?

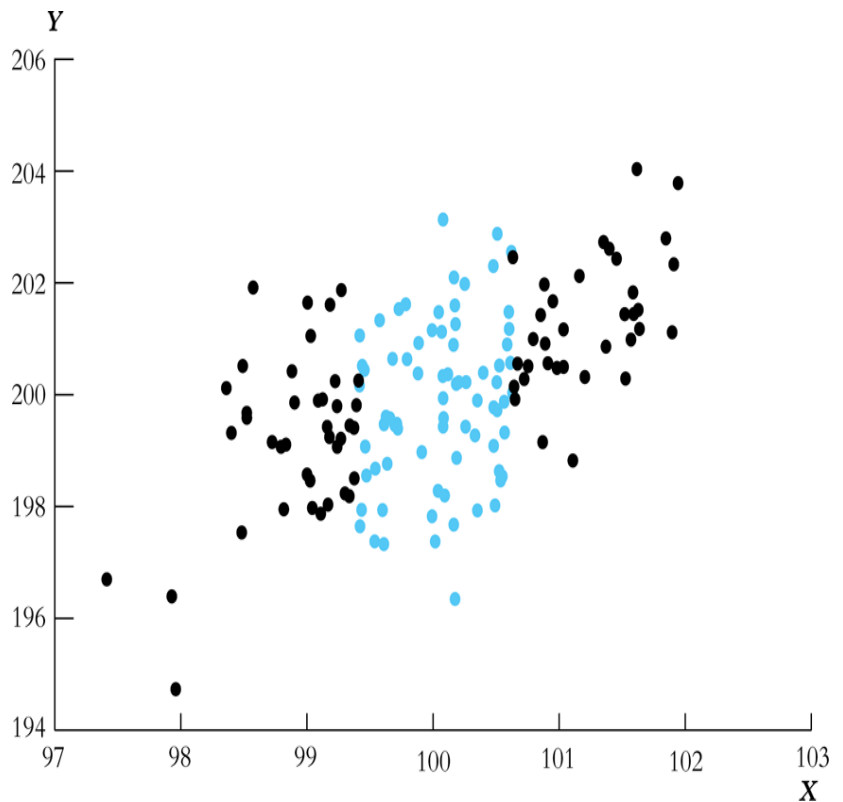




Valsts administrācijas
skola

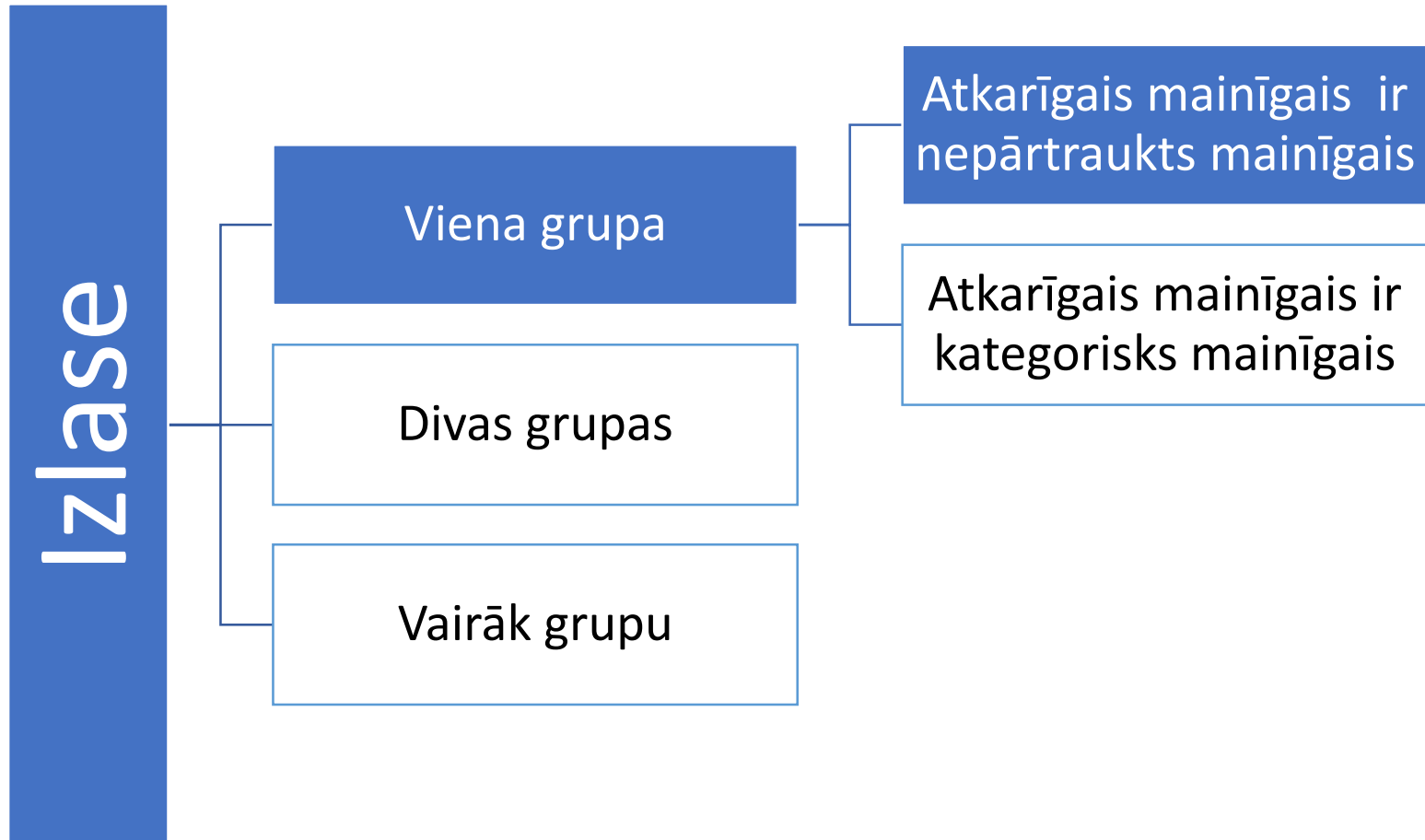
Neatkarīgo mainīgo analīze

- Melnās un zilās krāsas punktu skaits ir vienāds. Ar kuriem datiem var iegūt precīzāku regresijas taisni? Kura regresijas taisne precīzāk aprakstīs ģenerālkopu?





Vai MKM ir piemērotākā metode





Valsts administrācijas
skola

Vai vienādojums ir **teorētiski pamatots**?

Cik labi novērtētais vienādojums **atbilst datiem**?

Vai datu **izlase** ir pietiekami **liela un «tīra»**?

Vai dotam vienādojumam **MKM** ir **piemērotākā metode**?

Kā novērtētie parametri atbilst gaidām, kuras izveidojās pirms uzsāka datu vākšanu?

Vai **visi** acīmredzami **svarīgie mainīgie** ir **iekļauti** vienādojumā?

Vai ir izmantota teorētiski visloģiskākā **funkcijas forma**?

Vai regresija ir **brīva** no galvenajām **ekonometrijas problēmām**?



Modeļa specifikācija: neatkarīgie mainīgie un funkcijas veids

- Pēc atkarīgā mainīgā izvēles nākamais solis ir izveidot modeļa specifikāciju, kas nozīmē:
 - 1) neatkarīgo mainīgo izvēle un to mērvienības;
 - 2) mainīgo funkcionālā (matemātiskā) forma;
 - 3) kļūdu īpašības.
- Ja kādā no trīs soļiem pieļauj nepilnības, tad var izveidoties specifikācijas kļūdas:
 - 1) modelī iekļauti ne tikai teorētiski pamatoti izskaidrojošie mainīgie
 - 2) starp neatkarīgo un atkarīgo mainīgo sakarība ir acīmredzami nelineāra
 - 3) kļūdas neatbilst normālam sadalījumam
- Rezultātā, novērtētais modelis labi var aprakstīt izlases datus, bet lielas kļūdas rasties prognozējot ārpus izlases



Svarīgu mainīgo neiekļaušana un lieku mainīgo pievienošana

- Divi iemesli, kāpēc **svarīgi mainīgie nav pievienoti** modelim:
 - aizmirsts ...
 - dati nebija pieejami, to vākšanas laikā
- Grūti konstatēt, bet par to var liecināt:
 - iegūtie rezultāti ir ar negaidītām vērtībām
 - ir vairāki teorētiski loģiski mainīgie, kurus varētu pievienot modelim
- Par **nepiemērotu mainīgo iekļaušanu** modelī, var liecināt:
 - iepriekš statistiski nozīmīgi koeficienti, zaudē statistisko nozīmīgumu;
 - samazinās koriģētā determinācijas koeficienta vērtība



Alternatīvās funkciju formas

Funkcijas forma	Vienādojums	virziena koeficienta β_1 interpretācija
lineāra	$Y_i = \beta_0 + \beta_1 X_i + u_i$	par cik vienībām izmainās Y, ja X palielina par 1 vienību
log-log	$\ln Y_i = \beta_0 + \beta_1 \ln X_i + u_i$	par cik procentiem izmainās Y, ja X palielina par 1 %
lin-log	$Y_i = \beta_0 + \beta_1 \ln X_i + u_i$	par cik vienībām izmainās Y, ja X palielina par 1 %
log-lin	$\ln Y_i = \beta_0 + \beta_1 X_i + u_i$	par cik procentiem izmainās Y, ja X palielina par 1 vienību
polinoms	$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$	Aptuveni par cik vienībām izmainās Y, ja X palielina par vienu vienību un X vērtība ir tuvu nullei
inversa	$Y_i = \beta_0 + \beta_1 \left(\frac{1}{X_i} \right) + u_i$	Apgriesti proporcionālās Y izmaiņas, ja X palielina par vienu vienību un X vērtība ir tuvu nullei



Fiktīvie mainīgie

- Fiktīvie mainīgie pieņem vērtības 0 vai 1, atkarībā no kvalitatīva mainīgā nosacījumiem
 - piemēram, sieviete -> 1 , vīrietis -> 0 / pirmdiena -> 1 , cita nedēļas diena -> 0
- Vispārīgā veidā

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i, \text{ kur}$$
$$D_i = \begin{cases} 1, & \text{ja } i - \text{tam novērojumam izpildās nosacījums} \\ 0, & \text{citos gadījumos} \end{cases}$$



Valsts administrācijas
skola

Situācijas analīze

Datu failā «dati_ipasums» apkopotas 321 nekustamā īpašuma cenas un tos raksturojošie rādītāji: īpašuma vecums, ēkas un zemes platības, istabu skaits, vannasistabu skaits, palīgēku skaits, attālums līdz Rīgai

Izmantojiet dotos datus, statistikas programmu SPSS, atbildiet uz sekojošiem jautājumiem

- Regresijas modelī, kurā kā īpašuma cenu izskaidrojošie lielumi ir ēkas platība un istabu skaits, izvērtēt alternatīvās funkciju formas.
- Kura no tām varētu būt labākā ne tikai no izskaidrospējas viedokļa, bet arī labāk interpretējama?



Valsts administrācijas
skola

Vai vienādojums ir **teorētiski pamatots**?

Cik labi novērtētais vienādojums **atbilst datiem**?

Vai datu **izlase** ir pietiekami **liela un «tīra»**?

Vai dotam vienādojumam **MKM** ir **piemērotākā metode**?

Kā novērtētie parametri atbilst **gaidām**, kuras izveidojās pirms uzsāka datu vākšanu?

Vai **visi** acīmredzami **svarīgie mainīgie** ir **iekļauti** vienādojumā?

Vai ir izmantota teorētiski visloģiskākā **funkcijas forma**?

Vai regresija ir **brīva no galvenajām ekonometrijas problēmām**?



Regresiju analīzes klasiskie pieņēmumi

- Lai iegūtu labākos parametru novērtējumus ar MKM, jāizpildās septiņiem nosacījumiem
 1. Regresijas modelis ir lineārs, tā specifikācija ir pareiza, un kļūdas ir aditīvs lielums
 2. Kļūdu vidējā vērtība ir nulle
 3. Izskaidrojošie mainīgie nekorelē ar regresijas kļūdu
 4. Novērojumu kļūdas savā starpā nekorelē (nav autokorelācijas)
 5. Kļūdu dispersija ir konstanta (nav heteroskedastitātes)
 6. Neviens no izskaidrojošiem mainīgiem nav citu mainīgo lineāra kombinācija (nav multikolinearitātes)
 7. Regresijas kļūdu sadalījums atbilst normālam



3. Izskaidrojošie mainīgie nekorelē ar regresijas kļūdu

- Pieņemsim, ka dots regresijas modelis
 - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + u_i \quad (i = 1, 2, \dots, N)$
- Modelis ir:
 - lineārs (pēc koeficientiem)
 - kļūda ir aditīvs lielums
 - visi svarīgie izskaidrojošie mainīgie modelī ir iekļauti, t.i., modeļa specifikācija ir korekta, un tāpēc kļūdas ir stohastiskas ar nulles vidējo vērtību.
- Ja kaut viens no izskaidrojošiem mainīgajiem X_i korelē ar regresijas kļūdu, tad daļa no Y variācijām tiek izskaidrota ar kļūdas komponenti.
 - piemēram, ja regresijas kļūdas un X korelācija ir pozitīva, tad novērtētā koeficienta vērtība būs lielāka nekā patiesā



Valsts administrācijas
skola

3. Izskaidrojošie mainīgie nekorelē ar regresijas kļūdu

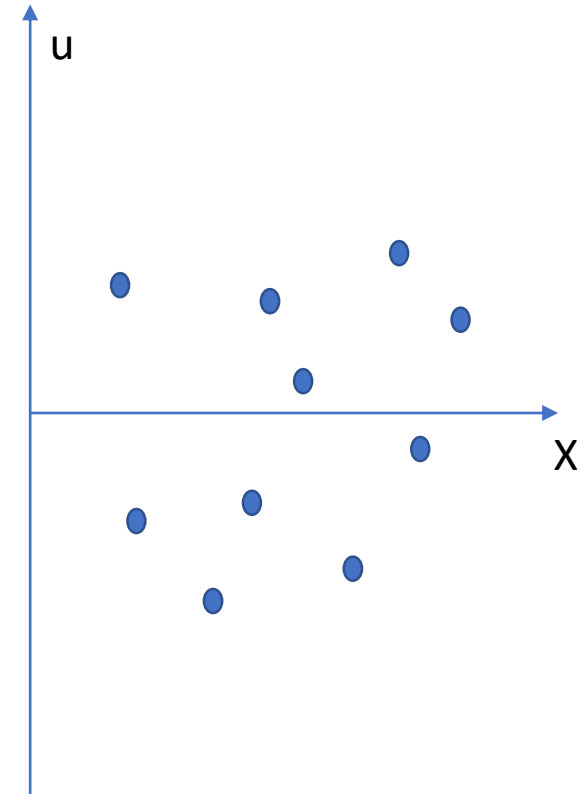
Kāpēc neizpildās:

tā bieži notiek, ja regresijā nav
iekļauts kāds ļoti svarīgs
izskaidrojošais mainīgais

Kā konstatatēt:

pietiekami aplūkot garafiski kļūdu
vērtības atkarībā no X

Vai attēls liecina par kļūdu un X
korelāciju?





4. Novērojumu kļūdas savā starpā nekorelē (nav autokorelācijas)

- Ja pastāv sistemātiska korelācija starp regresijas kļūdām, tad,
 - lai arī novērtētie koeficienti nav nobīdīti,
 - ir problemātiski ar MKM iegūt koeficientu standartkļūdu novērtējumus, un
 - pieaug iespēja kļūdaini pieņemt vai noraidīt hipotēzi par novērtētā koeficienta statistisko nozīmību.
- Ja secīgām X vērtībām atbilstošās regresijas kļūdas korelē, tad saka, ka pastāv kļūdu autokorelācija.

Autokorelācijas cēlonis:

regresijas vienādojumam izvēlēta neatbilstoša funkcionālā forma

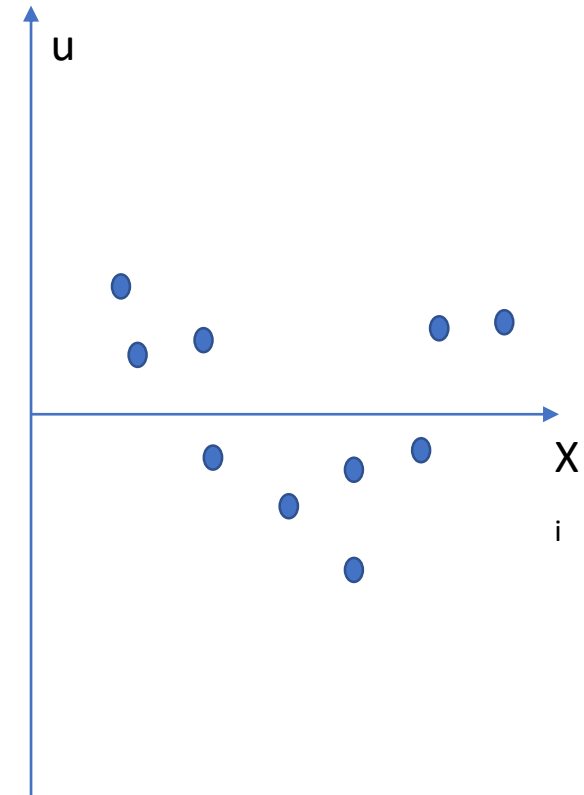
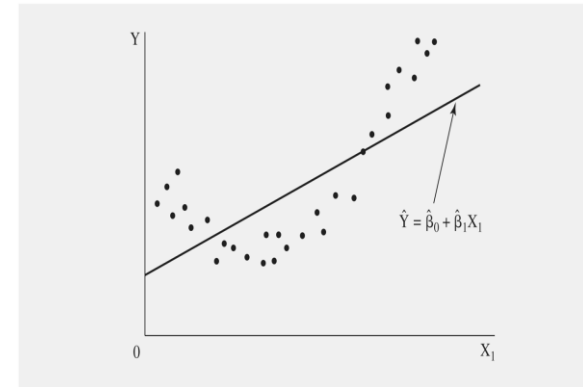


Valsts administrācijas
skola

4. Novērojumu kļūdas savā starpā nekorelē (nav autokorelācijas)

Kā konstatēt kļūdu
autokorelāciju:

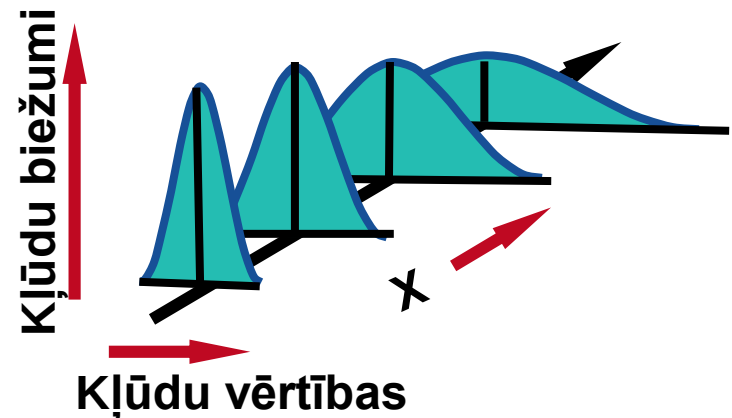
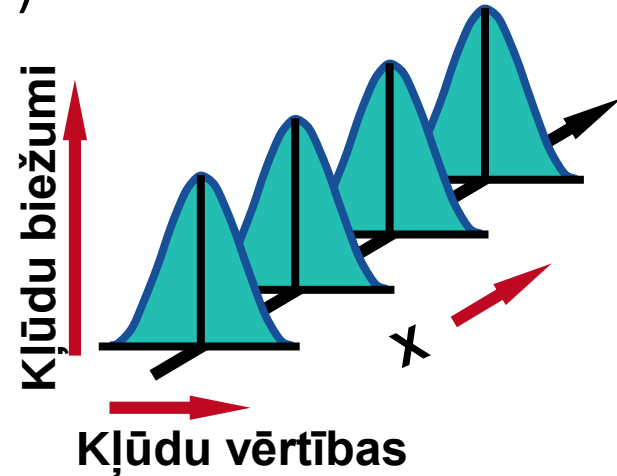
- ar Durbina – Vatsona
d- testa palīdzību
 - trīs pamatgadījumi:
 - izteikti pozitīva
autokorelācija
 $d > 0$
 - izteikti negatīva
autokorelācija
 $d < 0$
 - nav
autokorelācijas
problēmas $d \approx 0$
- grafiski





5. Kļūdu dispersija ir konstanta (nav heteroskedastitātes)

- Kļūdu dispersijai nevajadzētu mainīties, atkarībā no novērojumiem. Ja tas neizpildās, tad pastāv kļūdu heteroskedastitātes problēma
 - lai arī novērtētie koeficienti nav nobīdīti
 - un, tā kā parasti standartklūdas tiek novērtētas zemāk par patiesām,
 - pieaug iespēja kļūdaini noraidīt nulles hipotēzi, ka novērtētais koeficients nav statistiski nozīmīgs.
- Iemesls: modelim nav pievienots svarīgs izskaidrojošais mainīgais
 - nepievienotā mainīgā efektu, absorbē kļūdas





Valsts administrācijas
skola

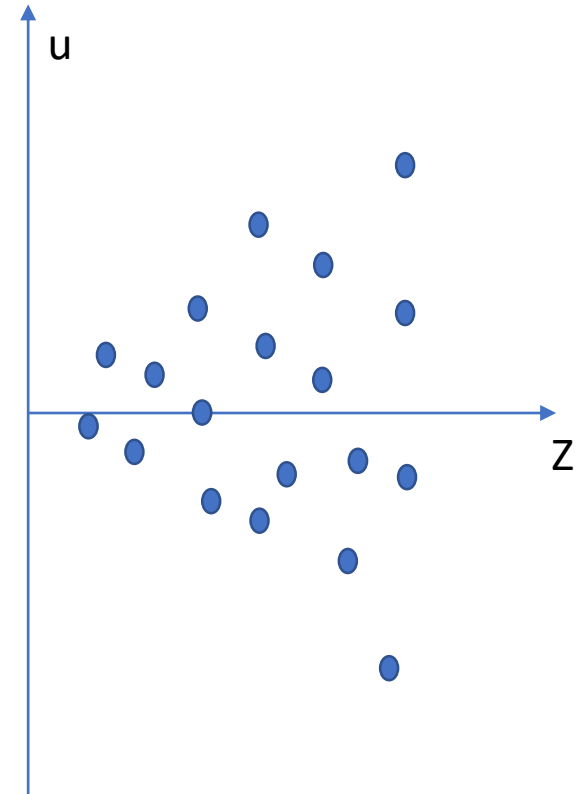
5. Kļūdu dispersija ir konstanta (nav heteroskedastitātes)

Kā konstatēt heteroskedastitātes problēmu:

- grafiski attēlojot kļūdas atkarībā no modelī neiekļautā izskaidrojošā mainīgā Z

Ar statistisko testu palīdzību:

- Pārka tests
- Vaita tests





Valsts administrācijas
skola

6. Neviena no izskaidrojošiem mainīgajiem nav citu mainīgo lineāra kombinācija (nav multikolinearitātes)

- Perfekta multikolinearitāte starp vairākiem neatkarīgajiem mainīgajiem nozīmē, ka
- vismaz viens mainīgais ir izsakāms kā summa no citu mainīgo reizinājuma ar konstantēm
- $X_k = a_1X_1 + \dots + a_iX_i$
- Perfektas multikolinearitātes gadījumā nav iespējams novērtēt modeļa koeficientus
 - statistiskās programmas parasti izdod kļūdas paziņojumu
 - risinājums – no modeļa jāizmet viens no kolineārajiem mainīgajiem



Multikolinearitātes sekas

- Ja multikolinearitāte nav perfekta, t.i., starp vairākiem neatkarīgiem mainīgiem pastāv pietiekami cieša lineāra sakarība, tad
 - lai arī novērtētie koeficienti nav nobīdīti,
 - to standartklūdas pieaug un t-vērtības samazinās, tāpēc
 - palielinās iespēja kļūdaini pieņemt nulles hipotēzi, ka novērtētais koeficients nav statistiski nozīmīgs
- Kā konstatēt multikolinearitāti
 - modelis ir ļoti jūtīgs pret izmaiņām, jaunu mainīgo pievienošana, vai kādu mainīgo atmešana, būtiski izmaina pārējo mainīgo koeficientu vērtības
 - starp mainīgajiem ir augsta korelācija
 - neatkarīgo mainīgo grupa, darbojoties kopā, var radīt multikolinearitātes problēmu, bez augstas pāru korelācijas
 - liela dispersijas inflācijas faktora vērtība (Variance Inflation Factors VIF)



Valsts administrācijas
skola

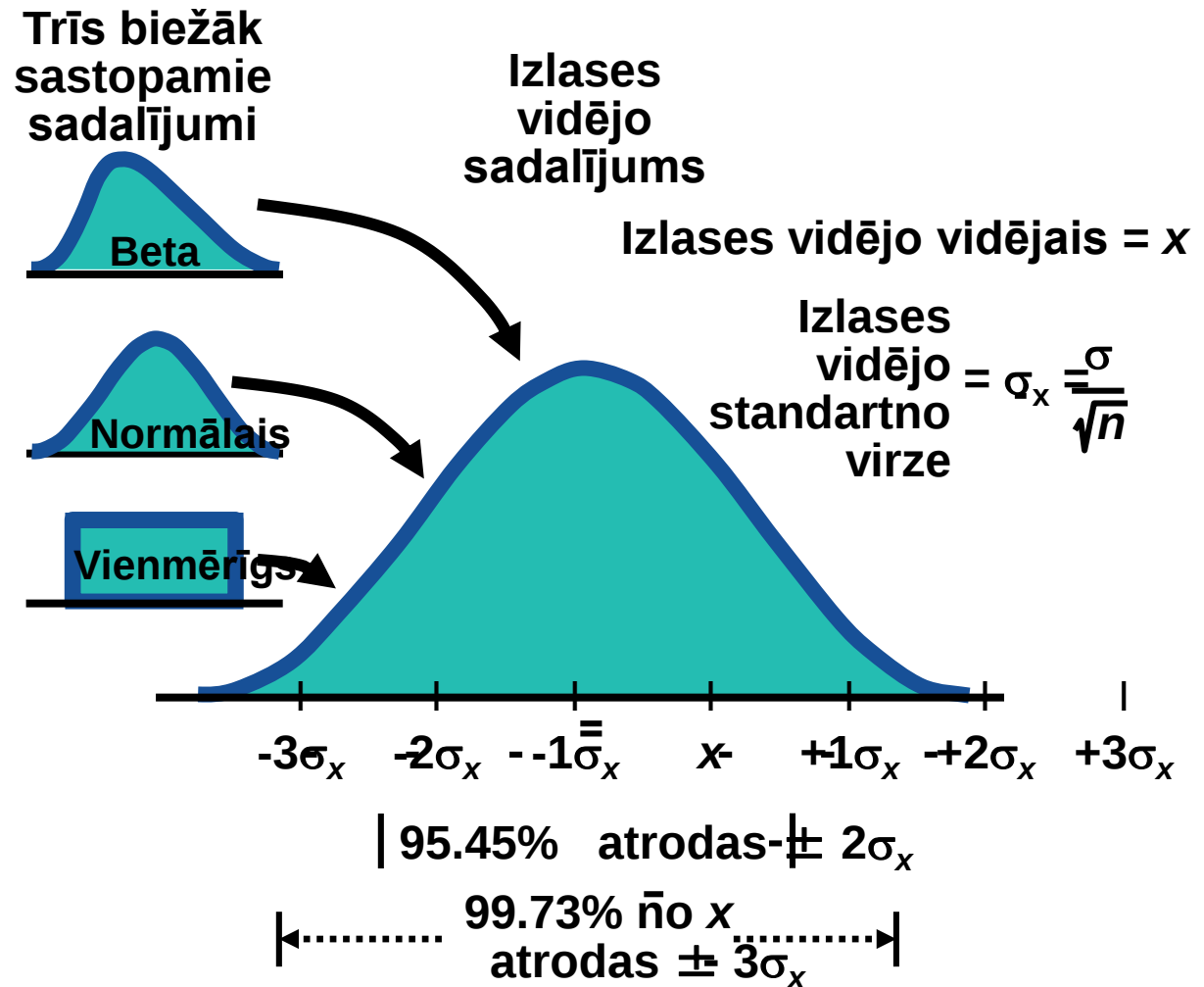
Multikolinearitātes novēršana

- Trīs risinājumi
 1. Nedarīt neko – atmetot mainīgo, kam vajadzētu būt modelī pēc teorijas, var radīt nepareizas modeļa specififikācijas problēmas
 2. Atmest vienu no izskaidrojošo mainīgiem, kas ir lienāri atkarīgs no citiem – lai izvēlētos kuru atmest, jābalstās uz teoriju
 3. Palielināt izlases apjomu – pieaugot izlasei, novērtēto koeficientu dispersija samazinās, un samazinās multikolinearitātes ietekme



7. Regresijas kļūdu sadalījums atbilst normālam

- Kļūda sadalījuma vidējā vērtība ir nulle
- un simetrisks pret vidējo
- Pielietojums: hipotēžu pārbaudei





Valsts administrācijas
skola

Situācijas analīze

Datu failā «dati_ipasums» apkopotas 321 nekustamā īpašuma cenas un tos raksturojošie rādītāji: īpašuma vecums, ēkas un zemes platības, istabu skaits, vannasistabu skaits, palīgēku skaits, attālums līdz Rīgai

Izmantojiet dotos datus, statistikas programmu SPSS, atbildiet uz sekojošiem jautājumiem

- Vai labākais no regresijas modeļiem, kurā kā īpašuma cenu izskaidrojošie lielumi ir ēkas platība un istabu skaits, ir brīvs no ekonometrijas pamatproblēmām: multikolinearitātes, autokorelācijas un heteroskedastitātes. Ja problēmas ir konstatētas, kā varētu tās novērst?



Valsts administrācijas
skola

Regresijas projekts ar SPSS



Regresijas projekta ar SPSS apraksts

- Datu failā «dati_uzkrajums» apkopotas 100 respondentu atbildes par mēneša ienākumiem un izdevumiem, kā arī papildus informācija: vecums, vai pilsonis, izglītība (kā izglītojoties pavadītais gadu skaits), cilvēku skaits mājsaimniecībā, uzkrājumiem pieejamā summa.
 - Izmantojiet dotos datus, statistikas programmu SPSS, izveidojiet modeli, kas izskaidrotu uzkrājumiem pieejamo summu.
 - Pārliecinieties, ka modelis atbilst ekonometrijas pamatpieņēmumiem!
- ☐ Vai vienādojums ir **teorētiski pamatots?** (diskusija)
 - ☐ Cik labi novērtētais vienādojums **atbilst datiem?**
 - ☐ Vai datu **izlase** ir pietiekami **liela un «tīra»?**
 - ☐ Vai dotam vienādojumam **MKM** ir **piemērotākā metode?**
 - ☐ Kā novērtētie parametri atbilst **gaidām**, kuras izveidojās pirms uzsāka datu vākšanu?
 - ☐ Vai **visi** acīmredzami **svarīgie mainīgie ir iekļauti** vienādojumā?
 - ☐ Vai ir izmantota teorētiski visloģiskākā **funkcijas forma?**
 - ☐ Vai regresija ir **brīva no galvenajām ekonometrijas problēmām?**
 - ☐ ir/nav iekļauti svarīgie mainīgie
 - ☐ multikolinearitāte
 - ☐ autokorelācija
 - ☐ heteroskedastitāte



Valsts administrācijas
skola

Paldies par Jūsu uzmanību!

Kontaktinformācija – Edgars Brēķis, e-pasts Edgars.Brekis@riseba.lv

Mācības finansētas no Eiropas Sociālā fonda projektiem

Nr.3.4.2.0/15/I/001 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide labāka regulējuma izstrādē mazo un vidējo komersantu atbalsta jomā" un

Nr.3.4.2.0/15/I/002 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide korupcijas novēršanas un ēnu ekonomikas mazināšanas jomā"

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA

Eiropas Sociālais
fonds

I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē