10 Paskaita. Dauginė tiesinė regresija.

10.1 Klasikinis tiesinis regresinis modelis kelių regresorių atveju

Daugeliu atveju yra daugiau nei vienas regresorius, turintis įtakos priklausomam kintamajam. Dauginės regresijos modelis apibūdina, kaip vienas atsako kintamasis tiesiškai priklauso nuo kelių iš anksto nustatytų kintamųjų. Dviejų regresorių modelį galime užrašyti taip:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon, \tag{1}$$

Pateiksime pavyzdį, kai yra 2 egzogeniniai kintamieji.
Pavyzdys. Nagrinėjamas ekonominis modelis yra toks:

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT, \tag{2}$$

čia $SALES$ yra mėnesio pajamos iš hamburgerių pardavimų, $PRICE$ yra prekės kaina duotame mieste, $ADVERT$ – reklamos išlaidos. Kintamieji $SALES$ ir $ADVERT$ matuojami tūkstančiais dolerių, o kintamasis $PRICE$ – doleriais.

Išsiaiškinsime nežinomų parametrų prasmę. Matematiškai laisvasis narys β_1 reiškia pajamas, kai kintamieji $PRICE$ ir $ADVERT$ įgyja nulines reikšmes. Taciau daugeliu atvejų šis koeficientas neturi aiškios ekonominės interpretacijos. Paprastai mes visuomet įtrauksime laisvąjį narį į modelius, nes tokie modeliai geriau aprašo ekonominius reiškinius ir duoda geresnes prognozes.

Kiti du parametrai reiškia priklausomo kintamojo pasikeitimą vienam nepriklausomam kintamajam pasikeitus vienu vienetu, o kitam nesikeičiant. Pavyzdžiui, β_2 rodo pajamų $SALES$ pasikeitimą tūkstančiais dolerių, kai kaina $PRICE$ pasikeičia vienu doleriu, o išlaidos reklamai nekinta. Kitaip tariant, β_2 yra kintamojo $SALES$ dalinė išvestinė kintamojo $PRICE$ atžvilgiu:

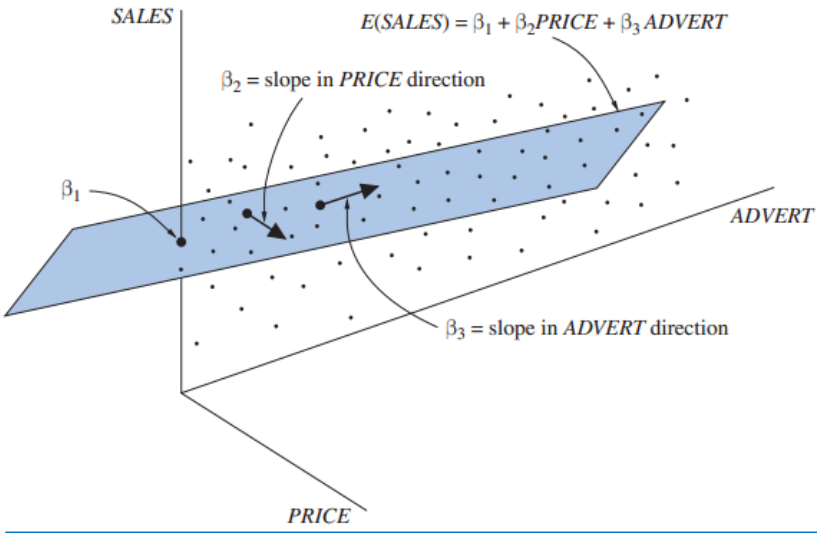
$$\beta_2 = \frac{\partial SALES}{\partial PRICE}.$$

Parametras β_2 gali būti tiek teigiamas, tiek neigiamas. Jei jis yra teigiamas, tai kainos padidėjimas sąlygoja pajamų padidėjimą, o tai reiškia, kad paklausa yra neelastiška kainai. Paklausos elastiškumas kainai reiškia, kad didėjant kainai, pajamos mažėja ($\beta_2 < 0$).

Parametras β_3 lygus pajamų pasikeitimui (\$1000), kai išlaidos reklamai $ADVERT$ padidėja vienu vienetu (\$1000), o kaina lieka pastovi:

$$\beta_3 = \frac{\partial SALES}{\partial ADVERT}.$$

Tikėtina, kad $\beta_3 > 0$, t.y. padidinus išlaidas reklamai, bendrosios pajamos padidės. Jei $\beta_3 < 1$, tai išlaidas reklamai padidinus \$1000, bendrosios pajamos



1 pav.: Regresijos plokštuma

padidės, bet mažiau nei \$1000, jei $\beta_3 > 1$, tai pajamos padidės daugiau nei \$1000. Tokiu būdu parametras β_3 parodo ir išlaidų reklamai atsiperkamumą. Taigi siekiant nustatyti reklamos politiką, labai svarbu teisingai įvertinti parametrą β_3.

Skirtingai nuo porinės tiesinės regresijos, kuri geometriškai vaizduojama tiese, dviejų regresorių (nepriklausomų kintamųjų) atveju modelis vaizduojamas plokštuma (1 pav.) β_1 yra taškas, kuriame plokštuma kerta $SALES$ ašį. β_2 yra kampas, kurį plokštuma sudaro su teigiama $PRICE$ ašies kryptimi, o β_3 – kampas su $ADVERT$ ašies teigiama kryptimi.

Modeliai su dviem prognozuojančiais kintamaisiais (pvz., x_2 ir x_3) ir atsako kintamuoju y gali būti suprantami kaip dvimatis paviršius erdvėje. Šio paviršiaus forma priklauso nuo modelio struktūros. Stebėjimai yra erdvės $\mathcal{R}^3$ taškai, o paviršius yra „parinktas“ taip, kad būtų geriausiai suderintas su stebėjimais. Jei egzogeniniai kintamieji į modelį įeina pirmaisiais laipsniais, regresijos paviršius bus plokštuma trimatėje erdvėje. Sudėtingesni modeliai gali apimti aukštesnius vieno ar daugiau egzogeninių kintamųjų laipsnių, pvz.,

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \varepsilon \tag{3}$$

arba

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{23} X_2 X_3 + \varepsilon. \tag{4}$$

Pastarieji du modeliai taip pat gali būti vadinami tiesiniais regresiniais modeliais, nes jie gali būti užrašyti kaip tiesiniai β-parametrų deriniai. (3) tipo modeliai kartais klaidingai vadinami netiesiniais regresiniais modeliais arba polinominės regresijos modeliais, nes regresijos kreivė nėra tiesė. (4) tipo modeliai

paprastai vadinami tiesiniais modeliais su kintamųjų sąveika. Jų paviršiai gali būti skirtingų formų, priklausomai nuo modelio parametrų reikšmių, kai lygio linijos yra lygiagrečios tiesės, parabolės arba elipsės.

Bendrasis tiesinis regresinis modelis, atitinkantis k regresorių, aprašomas lygtimi

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon, \tag{5}$$

čia Y – endogeninis kintamasis, (dar vadinamas regresantu arba paaiškinamuoju kintamuoju), $X_2, \ldots, X_k$ – egzogeniniai kintamieji (dažnai dar vadinami regresoriais arba paaiškinančiaisiais kintamaisiais), ε – atsitiktinis faktorius, $\beta_1, \ldots, \beta_k$ išreikštiniai modelio parametrai. Duomenys generuojami atsitiktinio proceso

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \varepsilon_t, t = 1, 2, \ldots, T. \tag{6}$$

Parametrai $\beta_1, \ldots, \beta_k$ atitinka regresorius $x_1, \ldots, x_k$, kiekvienas β_i parodo x_i įtaką kintamajam Y, kai visų kitų nepriklausomų kintamųjų reikšmės fiksuotos. Laisvasis narys β_1 atitinka x_1, kuris visada lygus 1. Lygtį (6) patogiau užrašyti matriciniu-vektoriniu būdu:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{7}$$

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}, \mathbf{E} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ \vdots & \vdots & \ddots & \\ 1 & X_{2T} & \ldots & X_{kT} \end{pmatrix}.$$

Matrica $\mathbf{X}$ vadinama projektine arba plano (*angl.* design) matrica. Atskiru atveju kai $k = 3$ šios matricos pavidalas:

$$\mathbf{X} = \begin{pmatrix} 1 & X_{21} & X_{31} \\ \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} \end{pmatrix}.$$

Reikalaujama, kad stebimų reikšmių būtų daugiau, negu modelio parametrų, t.y. $T > k$. Mūsų tikslas mažiausių kvadratų regresijoje yra parinkti hiperplokštumą $(k+1)$-matėje erdvėje, kuri minimizuotų liekanų kvadratų sumą.

10.2 Modelio prielaidos

Pirmosios dvi prielaidos yra egzogeniniams kintamiesiems.

EGZ1. Egzogeniniai kintamieji $X_2, \ldots, X_k$ nėra atsitiktiniai (stochastiniai) dydžiai. Manome, kad šių kintamųjų reikšmės mums žinomos *apiori*, t.y dar prieš kintamojo Y reikšmes. Pavyzdžiui, hamburgerių tinklo pavyzdyje mes iš anksto planuojame kainą ir investicijas į reklamą.

EGZ2. Ne vienas egzogeninis kintamasis nėra tiesinė kombinacija likusių egzogeninių kintamųjų, t.y. ne vienas iš jų nėra perteklinis kintamasis, o plano matrica $\mathbf{X}$ yra pilno rango matrica (nes jos eilutės nėra kolinearūs vektoriai).

Jei neišpildyta EGZ2 prielaida, tokia būsena vadinama *angl.* exact collinearity, mažiausių kvadratų procedūra negali būti taikoma.

Bendrojo tiesinio regresinio modelio (6) paklaidoms $\varepsilon_t, t = 1, 2, \ldots, T$ keliamos šios prielaidos:

MR1. Nulinių vidurkių: $E\varepsilon_i = 0, i = 1, 2, ..., T$. Ši sąlyga reiškia, kad visų modelio padarytų netikslumų, praleistų kintamųjų ir pan. vidutinė įtaka kintamajam Y yra nulinė.

MR2. Homoskedastiškumo: $var(\varepsilon_i) = \sigma^2, i = 1, 2, ..., T$. Homoskedastinis modelis reiškia, kad kiekviename stebėjime slypinčiai informacijai būdingas vienodas neapibrėžtumas.

MR3. Nekoreliuotų paklaidų: $cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t - E(\varepsilon_t))(\varepsilon_s - E(\varepsilon_s))$, kai $t \neq s$

MR4. Gausinių paklaidų: $\varepsilon_i, i = 1, 2, ..., T$ yra nepriklausomi atsitiktiniai dydžiai, turi vienodą normalųjį skirstinį su nuliniu vidurkiu ir dispersija σ^2:

$$\varepsilon_i \sim N(0, \sigma^2).$$

Gauso-Markovo teorema. Esant bendrojo tiesinio regresinio modelio prielaidoms EGZ1 – EGZ2 ir MR1 – MR3, mažiausių kvadratų įverčiai yra modelio parametrų geriausi tiesiniai nepaslinkti įverčiai (BLUE), *angl. best linear unbiased estimators*.

10.3 Parametrų įvertinimas

Nagrinėsime tik vieną iš galimų parametrų įvertinimo metodų – mažiausių kvadratų metodą. Vertinsime tiesinio modelio (7) parametrus $\boldsymbol{B}$. Ieškosime tokių parametrų reikšmių, kurioms nuokrypių nuo Y kvadratų suma būtų mažiausia. Kadangi $EY_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt}$,

$$(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) = \underset{(b_1, b_2, \ldots, b_k) \in R^k}{\arg\min} \sum_{t=1}^{T} (Y_t - b_1 - b_2 X_{2t} - \ldots - b_k X_{kt})^2.$$

Tai ir yra parametrų $(\beta_1, \beta_2, \ldots, \beta_k)$ *mažiausių kvadratų įvertis*. Tokį įvertį paprasčiau rasti pasinaudojus matricine išraiška. Apibrėšime funkciją

$$f(\boldsymbol{B}) = (\boldsymbol{Y} - \boldsymbol{XB})^T (\boldsymbol{Y} - \boldsymbol{XB}) = \sum_{t=1}^{T} (\boldsymbol{Y_t} - \boldsymbol{X_t B})^2, \boldsymbol{B} = (\beta_1, \beta_2, \ldots, \beta_k)^T \in R^k.$$

Modelio, apibrėžto lygtimi (7) mažiausių kvadratų įvertis yra

$$\hat{\beta} = \underset{\boldsymbol{B} \in R^k}{\arg\min} f(\boldsymbol{B})$$

Kitaip tariant, $f(\boldsymbol{B})$ įgyja minimalią reikšmę, kai $\boldsymbol{B} = \hat{\beta}$. Rasime funkcijos $f(\boldsymbol{B})$ minimalią reikšmę.

$$f(\boldsymbol{B}) = \boldsymbol{Y}^t \boldsymbol{Y} - 2\boldsymbol{B}^t \boldsymbol{X}^t \boldsymbol{Y} + \boldsymbol{B}^t \boldsymbol{X}^t \boldsymbol{XB}.$$

Suskaičiavę šios funkcijos išvestinę $\boldsymbol{B}$ atžvilgiu ir prilyginę ją nuliui, gausime

$$-2\boldsymbol{X}^t\boldsymbol{Y} + 2\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{B} = 0.$$

Tarkime, kad matricos $\boldsymbol{X}^t\boldsymbol{X}$ rangas lygus k, t.y. egzistuoja jos atvirkštinė matrica $(\boldsymbol{X}^t\boldsymbol{X})^{-1}$. Išsprendę pastarąją lygtį $\boldsymbol{B}$ atžvilgiu, gauname β MK įvertį

$$\hat{\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}. \tag{8}$$

Jei atvirkštinė matrica neegzistuoja, lygčių sistema gali būti išspręsta, tačiau sprendinys gali būti nevienintelis. Matricos $\boldsymbol{X}^t\boldsymbol{X}$ atvirkštinė matrica egzistuoja, jei matricos X stulpeliai nėra tiesiškai priklausomi. Tai reiškia, kad nė vienas stulpelis negali būti parašytas kaip tiesinis kitų stulpelių derinys. Vektorius $\hat{\boldsymbol{Y}}$ apskaičiuotas pagal tiesinės regresijos modelį gali būti išreikštas kaip

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y},$$

$$\text{čia } \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1})\boldsymbol{X}^t. \tag{9}$$

Matrica $\boldsymbol{H}$ vadinama **kepurine matrica** (*angl.* hat-matrix). Regresijos liekanos gali būti užrašytos tokiais būdais

$$\hat{\varepsilon}_t = Y - \hat{Y} = Y - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} = \boldsymbol{(I - H)Y}.$$

Klasikinio tiesinio regresinio modelio dispersiją σ^2 įvertiname pasinaudoję regresijos liekanomis. Paprastai naudojamas šis dispersijos įvertis:

$$\hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = \frac{1}{T-k}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}.$$

Žinoma, bendru atveju, šios formulės yra skirtos tik kompiuteriniams skaičiavimams. 2 pav. parodyta, kaip apskaičiuojami modelio parametrų didžiausio tikėtinumo įverčiai hamburgerių pardavimo pavyzdyje.

Matome, kad

$$\hat{\beta}_1 = 118.91, \hat{\beta}_2 = -7.908, \hat{\beta}_3 = 1.863$$

Taigi regresijos lygtis yra

$$\hat{SALES} = 118.91 - 7.908PRICE + 1.863ADVERT, R^2 = 0.4483.$$

$$(se) \quad (6.35) \quad (1.096) \qquad (0.683)$$

Remiantis pateikta informacija galime konstruoti intervalinius įverčius arba tikrinti hipotezes apie parametrų reikšmes. Pav. 3 pateiktos t-reikšmės ir p-reikšmės hipotezėms $H_0 : \beta_i = 0$ tikrinti su alternatyvomis $H_1 : \beta_i \neq 0, i = 1, 2, 3$.

Žvelgiant į įvertintą regresijos modelį, galime daryti tokias išvadas:

```
> X <- as.matrix(cbind(1,price,advert))
> head(X)
       price advert
[1,] 1  5.69    1.3
[2,] 1  6.49    2.9
[3,] 1  5.63    0.8
[4,] 1  6.22    0.7
[5,] 1  5.02    1.5
[6,] 1  6.41    1.3
> Y <- as.matrix(sales)
> head(Y)
     [,1]
[1,] 73.2
[2,] 71.8
[3,] 62.4
[4,] 67.4
[5,] 89.3
[6,] 70.3
> beta_hat <- solve(t(X)%*%X)%*%t(X)%*%Y
> View(beta_hat)
> beta_hat
              [,1]
       118.913610
price   -7.907854
advert   1.862584
```

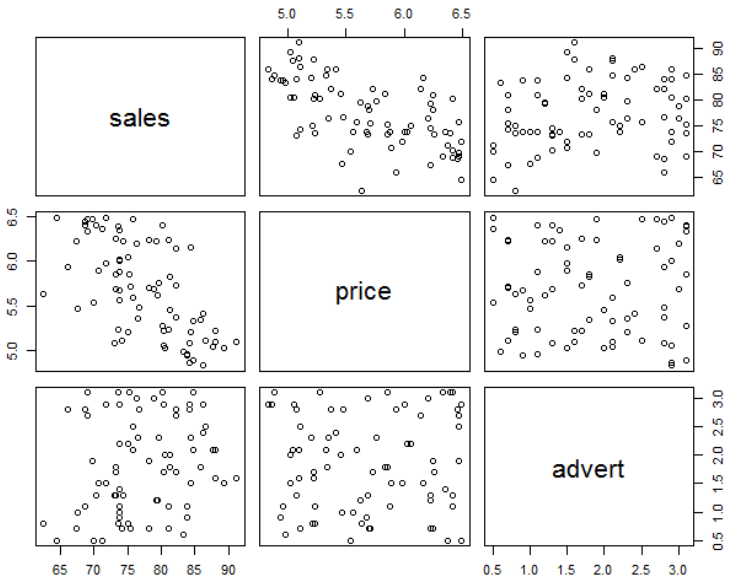
2 pav.: Regresijos koeficientų skaičiavimas.

```
> model <- lm(sales ~ price + advert)
> summary(model)

Call:
lm(formula = sales ~ price + advert)

Residuals:
     Min       1Q   Median       3Q      Max
-13.4825  -3.1434  -0.3456   2.8754  11.3049

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.9136     6.3516  18.722  < 2e-16 ***
price        -7.9079     1.0960  -7.215 4.42e-10 ***
advert        1.8626     0.6832   2.726  0.00804 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.886 on 72 degrees of freedom
Multiple R-squared:  0.4483,    Adjusted R-squared:  0.4329
F-statistic: 29.25 on 2 and 72 DF,  p-value: 5.041e-10
```

3 pav.: Dauginės regresijos modelis R.

```
> anova(model)
Analysis of Variance Table

Response: sales
          Df  Sum Sq Mean Sq F value    Pr(>F)
price      1 1219.09 1219.09 51.0631 5.946e-10 ***
advert     1  177.45  177.45  7.4326  0.008038 **
Residuals 72 1718.94   23.87
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4 pav.: ANOVA lentelė R.



5 pav.: Kintamųjų sklaidos diagramos.

```
> cor(andy[c(1,2, 3)])
             sales       price     advert
sales   1.0000000 -0.62554053 0.22208038
price  -0.6255405  1.00000000 0.02636585
advert  0.2220804  0.02636585 1.00000000
```

6 pav.: Kintamųjų koreliacijos.

```
> confint(model)
                2.5 %     97.5 %
(Intercept) 106.251852 131.575368
price       -10.092676  -5.723032
advert        0.500659   3.224510
```

7 pav.: Kintamųjų 95% pasikliautinieji intervalai.

1. Neigiamas koeficientas prie kainos (PRICE) rodo, kad paklausa elastinga kainai ir kainos padidėjimas vienu doleriu esant nepakitusiom reklamos išlaidoms bendras pajamas vidutiniškai sumažintų $7908. Arba, sumažinę kainą vienu doleriu, padidintume pajamas vidutiniškai $7908. Taigi kainos mažinimas, siūlant įvairias nuolaidas, padidintų pajamas. Kainos pasikeitimas $1 yra palyginti didelis kainos šuolis. Kainos imties vidurkis yra 5.69, standartinis nuokrypis yra 0.52. Realistiškiau nagrinėti 10 centų kainos pasikeitimą. Tokiu atveju pajamos pasikeistų $791.

2. Reklamos (ADVERT) koeficientas yra teigiamas. Vadinasi, esant pastoviai kainai, padidinus reklamos išlaidas $1000, pajamos padidėtų vidutiniškai $1863. Taigi reklamos išlaidų padidėjimas padidins pelną.

3. Laisvojo nario įverčio tiesioginė interpretacija – jei ir kaina, ir reklamos išlaidos būtų nulinės, pardavimo pajamos būtų $118 914. Akivaizdu, kad taip negali būti. Tai patvirtina teiginį, kad regresinis modelis gerai aprašo tikrovę tik toje duomenų kitimo srityje iš kurios mes turime imties duomenis. Šiame modelyje laisvasis narys vaidina modelio stabilizatoriaus vaidmenį ir pagerina prognozavimo tikslumą.

Regresijos lygtį galime naudoti pajamų prognozavimui. Tarkime, kad mėsainių tinklas nori numatyti pajamas, kai kaina yra $5.50, o išlaidos reklamai – $1200. Kadangi modelyje išlaidos reklamai ir pajamos skaičiuojamos tūkstančiais dolerių,

$$SALES = 118.91 - 7.908 \cdot 5.5 + 1.863 \cdot 1.2 = 77.656.$$

Esant minėtoms sąlygoms pelnas būtų $77 656.

Pastaba. Interpretuojant rezultatus tai reikia daryti atsargiai. Neigiamas ženklas prie kainos kintamojo reiškia, kad kuo mažesnė kaina, tuo didesnis pelnas. Remiantis tokia logika, mes turėtume kainą sumažinti iki nulio. Tačiau akivaizdu, kad tai nepadidins pardavimų. Iš to išplaukia labai svarbi išvada: įvertintas regresijos modelis aprašo ryšį tarp ekonominių kintamųjų dydžių tik srityje, artimoje imtyje esantiems duomenis. Duomenų ekstrapoliavimas link ekstremalių verčių paprastai nėra gera idėja, nes drastiškai padidėja paklaidos.

Liko įvertinti liekamojo nario dispersiją:

$$\sigma^2 = var(\varepsilon_i) = E(\varepsilon_i^2).$$

ε_i reikšmės nežinomos, o stebimos tik modelio liekanos

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3}).$$

Kadangi modelio parametrų skaičius $k = 3$, nepaslinktas σ^2 įvertis yra

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{T} \hat{\varepsilon}_i^2}{T-3} = \frac{1718.943}{75-3} = 23.874.$$

$$\hat{\sigma} = \sqrt{23.874} = 4.8861.$$

Pav. 4–5 pateiktos dydžių $SSE = \sum_{i=1}^{T} \hat{\varepsilon}_i^2 = 1718.94$ ir $\hat{\sigma} = 4.8861$ reikšmės, apskaičiuotos R. $\hat{\sigma}$ kartais yra žymimas kaip *Residual standard error* arba *root MSE* (mean squared error).

10.4 Multikolinearumo problema

Dauginė regresija geriausiai tinka prognozėms tada, kai visi nepriklausomi kintamieji $X_1, \ldots, X_n$ tarpusavyje nekoreliuoja, o priklausomybė sieja juos su Y. Kai tarp kintamųjų $X_1, \ldots, X_n$ yra stipriai koreliuojančių, tai yra taip vadinama **multikolinearumo problema**. Dėl multikolinearumo gali atsirasti "ne tas" daugiklio ženklas. Dar blogiau yra regresijos koeficientų nestabilumas, keli papildomi stebėjimai gali juos visiškai pakeisti, todėl toks modelis netinka prognozėms. Multikolinearumas nustatomas skaičiuojant taip vadinamus **dispersijos mažėjimo daugiklius** VIF (*angl.* variance inflation index). Tarkime, kad R_j^2 yra regresijos modelio, kuriame X_j yra priklausomas kintamasis, o $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$ – nepriklausomi kintamieji, determinacijos koeficientas. Tuomet kintamojo X_j dispersijos mažėjimo daugiklis

$$VIF_j = \frac{1}{1 - R_j^2}.$$

Paprastai VIF_j interpretuojamas kaip $\hat{b}_j$ dispersijos santykis su ta dispersija, kurią $\hat{b}_j$ turėtų, jei X_j nekoreliuotų su likusiais X. MX_j multikolinearumą galima įtarti, kai $VIF_j > 4$. Žiūrint tik į kintamųjų $X_1, \ldots, X_n$ koreliacijas multikolinearumą ne visada galima pastebėti.

Ką daryti aptikus multikolinearumą? Siūloma padidinti imtį. Jei tai nepadeda, atsisakyti dalies kintamųjų arba imti jų tiesinį darinį.