0.1 Увод у курс

"Essentially, all models are wrong, but some are useful"
George E.P. Box

Како се променом једне или више независних случајних промењливих мења вредност зависне случајне величине? Како одредити аналитичко-математички облик одговарајуће везе? Одговор на ова, као и на низ других питања даје нам управо регресија. Овај курс биће посвећен линеарној регресији. Идеје које се овде користе могу послужити и приликом анализирања других типова регресије.

Први записи о методи најмањих квадрата могу се наћи у радовима Лежандра и Гауса, почетком 19. века. Они су овај метод користили за одређивање орбита небеских тела око Сунца. Са речју "регресија" математичари су се први пут сусрели у раду Ф. Галтона , *Regression toward mediocrity in hereditary stature* из 1855. године. Он је дошао до закључка да синови веома високих очева нису тако високи. Иако је Галтон разлог за то пронашао у генетици, његов пример иницирао је проучавање ове теме од стране статистичара и тако почиње развој ове веома значајне статистичке области.

Дефиниција 0.1.1. *Регресија је зависност једне случајне промењливе од друге (или више њих). Регресиони модел је математички модел који описује ту зависност.*

Дефиниција 0.1.2. *Случајна величина $f(X) = E(Y|X)$ назива се регресиона функција, при чему X може бити вишедимензиона случајна величина.*

Следећа теорема оправдава облик функције регресије.

Теорема 0.1.1 (2).

$$E(Y - E(Y|X))^2 \leq E(Y - g(X))^2$$

за сваку функцију $g(X)$, уз претпоставку да постоји математичко очкевиање на десној страни неједнакости.

Доказ. □

Регресиона функција је права линија акко случајни вектор $(X, Y)^T$ има вишедимензионална нормалну расподелу. Регресиону праву има смисла конструисати и када знамо да заједничка расподела није нормална. Тада је то права која од свих правих линија најбоље описује зависност између Y и X у смислу средњеквадратног одступања.

Регресиони модел се може предцавити у облику

$$Y = f(X) + \varepsilon,$$

где је ε случајна промењлива независна од X, најчешће са нормалном $\mathcal{N}(0, \sigma^2)$ расподелом.

Уколико из нпр. графичког приказа зависности (X, Y) имамо разлога да претпоставимо да је $f(X) = aX + b$ онда се коефицијенти a, b одређују тако да се минимизира $E(Y - (aX + b))^2$.

Добија се да је

$$a = \frac{EXY - EXEY}{DX}$$
$$b = EY - aEX,$$

па се коефицијенти a, b могу оценити методом замене, односно

$$\hat{a} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\bar{S}_X^2} = \hat{\rho}\frac{\bar{S}_X}{\bar{S}_Y}$$
$$\hat{b} = \bar{Y} - \hat{a}\bar{X},$$

Уколико претпоставимо да X није случајна промењлива говоримо о *контролисаној регресији.*

Имајући у циду саму дефиницију регресионе функције, од сада па надаље можемо претпоставити да се ради о контролисаној регресији.

Да поновимо, главни наш задатак у овом курсу је да одговоримо на следећа питања.

Неколико питања на која ћемо дати одговор:

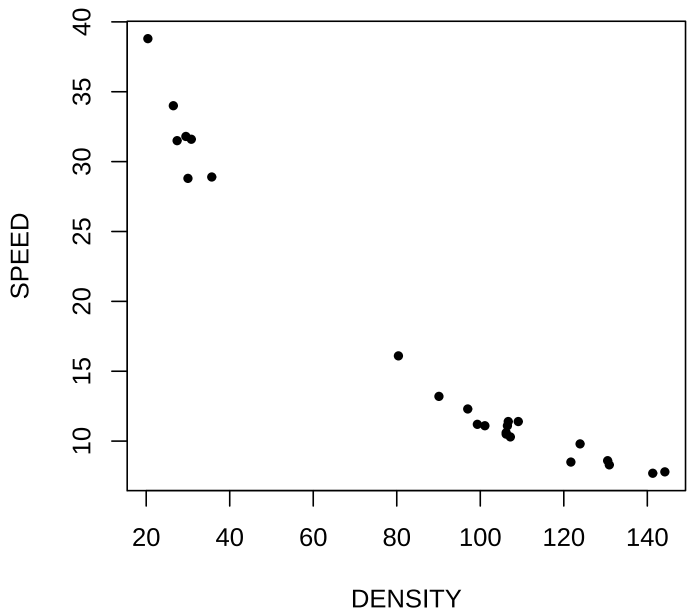
- Каква је веза између различитих обележја?
- Када одередимо облик модела како да оценимо његове параметре?
- Који су модели "допустиви" и у ком смислу?
- Како да испитамо квалитет модела?

Како изгледа цео процес бирања модела демонстрираћемо на наведеном примеру.

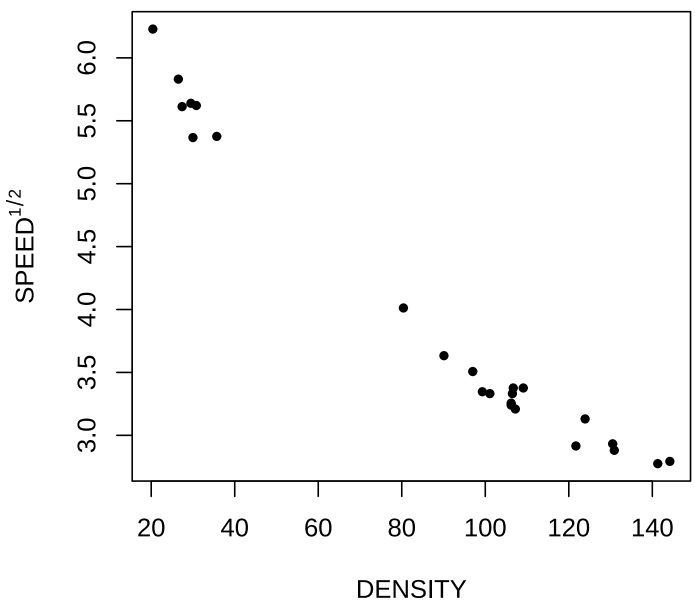
Пример 0.1.1. *У циљу истраживања у којој мери број возила на путу утиче на брзину возила сакупљани су подаци о "густини" возила (број аутомобила у једној миљи) и просечној брзини аутомобила.*

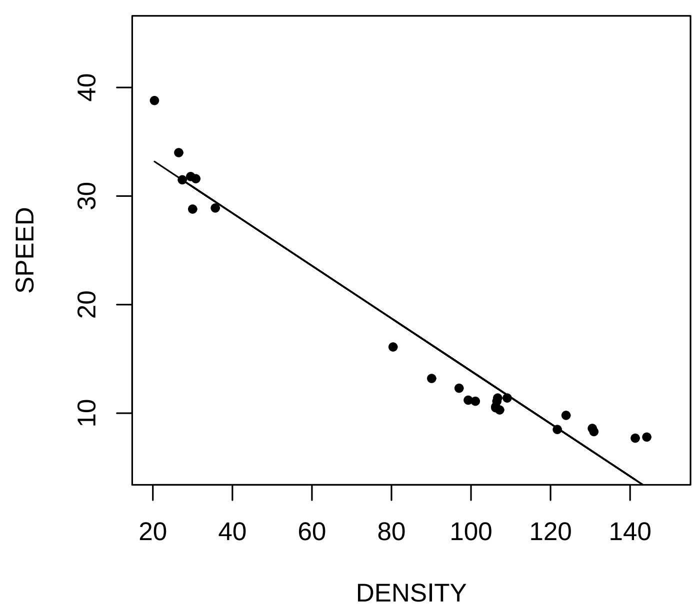
	DENSITY	SPEED
1	20.40	38.80
2	27.40	31.50
3	106.20	10.60
4	80.40	16.10
5	141.30	7.70
6	130.90	8.30
7	121.70	8.50
8	106.50	11.10
9	130.50	8.60
10	101.10	11.10
11	123.90	9.80
12	144.20	7.80
13	29.50	31.80
14	30.80	31.60
15	26.50	34.00
16	35.70	28.90
17	30.00	28.80
18	106.20	10.50
19	97.00	12.30
20	90.10	13.20
21	106.70	11.40
22	99.30	11.20
23	107.20	10.30
24	109.10	11.40

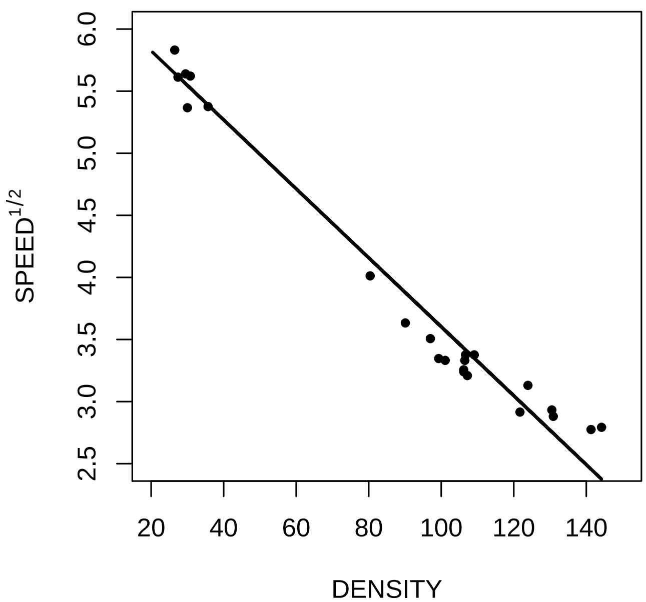
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	38.1295	1.2177	31.31	0.0000
E1.1$DENSITY	-0.2425	0.0126	-19.22	0.0000

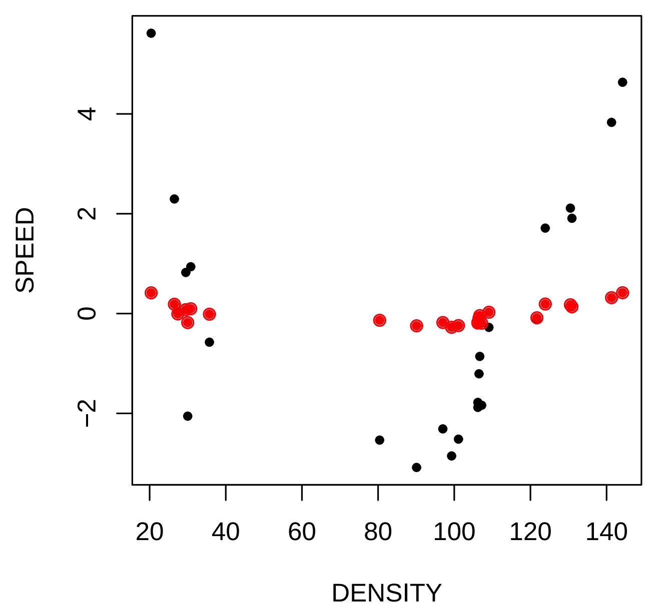


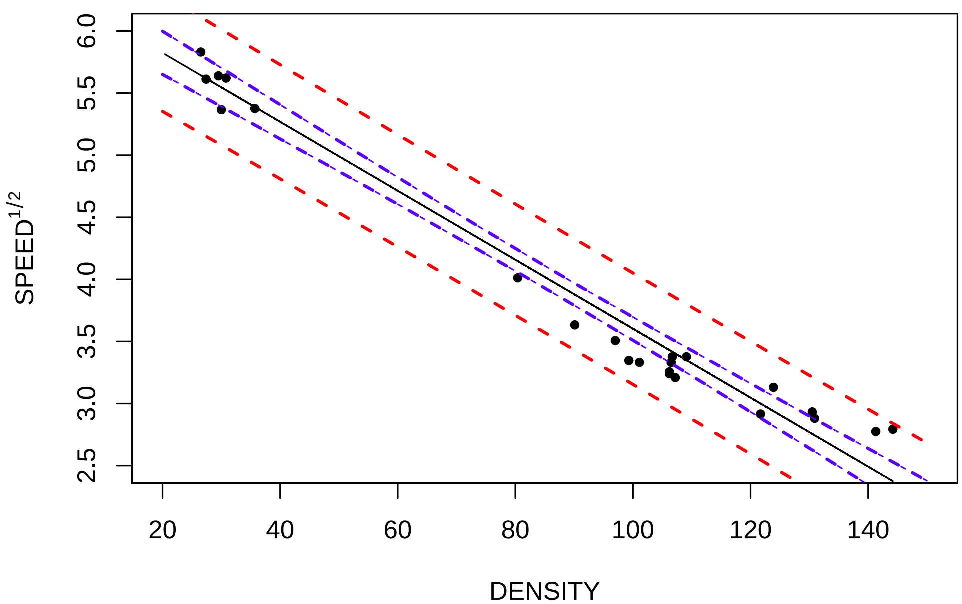
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.3797	0.1028	62.09	0.0000
E1.1$DENSITY	-0.0278	0.0011	-26.09	0.0000


SPEED$^{1/2}$
3.0
3.5
4.0
4.5
5.0
5.5
6.0
20
40
60
80
100
120
140
DENSITY


SPEED
10
20
30
40
20
40
60
80
100
120
140
DENSITY


SPEED^1/2
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
20
40
60
80
100
120
140
DENSITY


SPEED
−2
0
2
4
20
40
60
80
100
120
140
DENSITY


SPEED$^{1/2}$
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
20
40
60
80
100
120
140
DENSITY

Први корак је свакако да графички представимо податке и да уочимо неки облик зависности (ако постоји).

Са првог графика можемо закључити да са повећањем густине саобраћаја опада брзина истог, што је сасвим очекиван закључак. Најједноставнији модел који би могао да опише податке је линеарна веза, односно $y = ax + b + \varepsilon$, где је y просечна брзина аутомобила а x густина собраћаја. Јасно је да у модел морамо да укључимо и неки "шум" (ε) који би оправдао то што тачке на графику нису све колинеарне. Неке природне особине које тај шум треба да задовољава је да је 'мали', да је "центриран" око нуле, да не зависи од x и y итд. Шум заправо представља грешку модела.

Један од најпопуларнијих, најједноставнијих и слободно можемо рећи основних метода за оцену параметара модела је метод најмањих квадрата. Идеја је да параметре оценимо оним вредностима који минимизирају суму квадратних одступања оцењене од праве вредности, односно

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2.$$

Добијамо да су тражени $\hat{a}$ и $\hat{b}$

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

Приметимо да полазни модел можемо написати у центрираном облику $y_i = a(x_i - \bar{x}) + b + a\bar{x} + \varepsilon$. Испоставља се да је овај облик погоднији за прогнозирање јер $\hat{y}_i = \hat{a}(x_i - \bar{x}) + \bar{y}$.

Још је важно да се примети да је $\sum_{i=1}^{n} \varepsilon_i = 0$.

У (0.1.1) једначине правих које се добијају у првом, односно другом моделу су $y = -0.24x + 38.13$, односно $y = -0.028x + 6.38$.

Када је модел добар одступања оцењених вредности од правих (резидуали) су мали. Зато је природно за меру квалитета модела узети, за початак, $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$. Главни проблем са овом мером одступања је зато што она зависи од јединице. Зато ћемо искористи сличну идеју као за увођење коефицијента корелације.

$$SSTO = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$
$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = SSE + SSR$$

SSR је одступање које је објашњено моделом. Зато уводимо *коефицијент детерминације* R^2 као меру квалитета модела.

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

Јасно је да ако бисмо имали перфектан модел онда би $R^2 = 1$.

У примеру 0.1.1 за први модел се добија да је $R^2 = 0.94$ док је за други модел $R^2 = 0.98$.

Може се показати да је $R = |\rho_{xy}|$.

Напомена: не треба увек (само) користити R^2 као меру квалитета модела. О томе ће бити више речи у остатку курса.

0.2 Основни појмови из линеарне алгебре

Дефиниција 0.2.1. *Векторски простор V је непразан скуп, затворен за сабирање и скаларно множење. Његови елементи називају се вектори.*

Дефиниција 0.2.2. *Вектори $v_1, ..., v_n$ су линеарно независни ако не постоји нетривијално решење једначине*

$$x_1 v_1 + \cdots x_n v_n = 0.$$

Дефиниција 0.2.3. *Вектори v_1 и v_2 су ортогонални ако је $< v_1, v_2 >= 0$.*

Приметимо да из ортогоналности следи линеарна независност а да обрнуто не важи. На пример, посматрајмо векторе $v_1 = (1, 0, 1)^T$ и $v_2 = (0, 1, 1)^T$.

Дефиниција 0.2.4. *Квадратна матрица M је ортогонална ако је $M^T M = I$*

Дефиниција 0.2.5. *Ранг $n \times p$ матрице A у ознаци $R(A)$ је максималан број линеарно независних колона(врста).*

Из саме дефиниције следи следећи низ једнакости

$$R(A^T A) = R(AA^T) = R(A) = R(A^T).$$

Дефиниција 0.2.6. *Ако је за неко $\lambda \in R$, $Ax = \lambda x$ онда се λ назива сопствена вредност матрице A, а x сопствени вектор.*

Познато је да је за квадратну матрицу A је λ решење једначине $det(A - \lambda I) = 0$ и тада је $detA = \prod \lambda_i$.

Дефиниција 0.2.7. *Траг квадратне матрице $A = [a_{ij}]$ представља збир елемената на дијагонали, односно*

$$trA = \sum_i a_{ii}.$$

За траг матрице A важе следеће једнакости (под претпоставком да је множење дефинисано):

1. $tr(A + B) = tr(A) + tr(B)$;
2. Ако је A $n \times n$ матрица и P несингуларна матрица, онда је $tr(P^{-1}AP) = tr(A)$;
3. Ако је A $n \times n$ матрица и M ортогонална матрица, онда је $tr(M^T AM) = tr(A)$.
4. $tr(ABC) = tr(BCA)$

Дефиниција 0.2.8. *Квадратна матрица A је симетрична ако је $A = A^T$.*

Особине симетричне $n \times n$ матрице A:

1. Постоји ортогонална матрица $C = (c_1, ...c_n)$ и дијагонална матрица $\Lambda = (\lambda_1, ...\lambda_n)$ (сопствене вредности матрице) таква да је $A = C\Lambda C^T$ (спектрална декомпозиција матрице A). Тада је $A = \sum_{i=1}^n \lambda_i c_i c_i^T$.
2. $R(A)$ је број сопствених вредности различитих од нуле;
3. $tr(A) = \sum_{i=1}^n \lambda_i$;
4. Ако је A несингуларна матрица онда је $tr(A^{-1}) = \sum_{i=1}^n \lambda_i^{-1}$;

5. Постоји ортогонална трансформација $y = M^T x$ тако да је

$$x^T A x = \sum \lambda_i y_i^2;$$

6. $R(A+B) \leq R(A) + R(B)$

У случају несиметричне $n \times p$ матрице A матрица ранга r постоји декомпозиција $A = ULV$ где је $U^T U = I = V^T V = I$ и L је несингуларна матрица ранга r.

Дефиниција 0.2.9. *Нека је A симетрична матрица. Тада је са $Q(x) = x^T A x$ дефинисана једна квадратна форма вектора x. Кажемо да је Q позитивно (негативно) дефинитна ако за свако $x > 0$, $Q(x) > 0$ ($Q(x) < 0$). Ако се допушта једнакост онда је позитивно семи-дефинитна (негативно семи-дефинитна).*

Може се показати да ако је $A_{p \times p}$ позитивно дефинитна и $B_{k \times p}$ матрица ранга $k \leq p$ онда је BAB^T позитивно дефинитна.

Важи и следеће: Симетрична матрица A је позитивно дефинитна акко постоји несингуларна матрица P таква да је $A = P^T P$

Важи и следеће

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_{\max}, \tag{1}$$

$$\min_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_{\min}. \tag{2}$$

Дефиниција 0.2.10. *За матрицу P за коју је $P^2 = P$ кажемо да је идемпотента. Уколико је и симетрична онда се назива матрицом пројекције или пројектором.*

Особине пројектора:

1. $tr(P) = R(P)$;

2. P је позитивно семи дефинитна;

3. Нека су P_1 и P_2 пројектори. Ако је $P_1 - P_2$ позитивно семи-дефинитна онда је и пројектор, као и $P_1 P_2 = P_2 P_1 = P_2$

Доказ треће особине:

Из позитивне семидефинитности $P_1 - P_2$ закључујемо да је

$$((I-P_1)y)^T(P_1-P_2)((I-P_1)y) \geq 0.$$

Даље је $y^T(I-P_1)^T P_1(I-P_1)y = 0$, па је

$$((I-P_1)y)^T P_2((I-P_1)y) \leq 0.$$

Узимајући у обзир позитивну семидефинитност P_2 је

$$((I-P_1)y)^T P_2((I-P_1)y) = 0,$$

односно

$$((I-P_1))^T P_2((I-P_1)) = 0 = (I-P_1))^T P_2 P_2((I-P_1)).$$

Одавде је

$$P_2(I-P_1) = 0.$$

Дефиниција 0.2.11. *Матрица A максималног ранга има инверз A^{-1}. Уколико матрица није максималног ранга онда постоји уопштени инверз A^- за који важи $AA^-A = A$. Овај инверз не мора бити јединствен.*

0.2.1 Матрично диференцирање

Дефиниција 0.2.12. *Нека је X $n \times p$ матрица и f скаларна функција. Тада је матрично диференцирање дефинисано са*

$$\frac{\partial f(X)}{\partial X} := \left(\frac{\partial f(X)}{\partial x_{ij}}\right)$$

1. $\frac{\partial a^T x}{\partial x} = a$;
2. $\frac{\partial x^T x}{\partial x} = 2x$;
3. $\frac{\partial x^T A x}{\partial x} = (A + A^T)x$;
4. $\frac{\partial x^T A y}{\partial x} = Ay$;

0.3 Неке важне вишедименизионалне расподеле

0.3.1 Нормална расподела

Дефиниција 0.3.1. *Случајни вектор X има n-димензиона нормалну расподела $N_n(\mu, \Sigma)$ уколико је његова фунцкија густине*

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}}, \ \ x \in R^n$$

где је Σ симетрична, позитивно дефинитна коваријациона матрица а са $|\Sigma|$ је означена њена детерминанта.

Вишедимензионална нормална расподела има следеће лепе особине:

- Уколико X има $N(\mu, \Sigma)$ распделу онда $AX+b$ има $N(A\mu+b, A\Sigma A^T)$;
- Ако случајни вектор $Z = (X^T, Y^T)^T$ има $N\left(\left(\begin{array}{c} \mu_x \\ \mu_y \end{array}\right), \left(\begin{array}{cc} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{array}\right)\right)$ расподелу

  онда су маргиналне расподеле за X и Y редом $N(\mu_x, \Sigma_x)$ и $N(\mu_y, \Sigma_y)$, а условне расподеле

  $X|y \sim N(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{xy}^T)$

  $Y|x \sim N(\mu_y + \Sigma_{xy}^T\Sigma_{xx}^{-1}(x - \mu_x), \Sigma_y - \Sigma_{xy}^T\Sigma_x^{-1}\Sigma_{xy})$;
- Момент генераторна функција случајног вектора X са $N(\mu, \Sigma)$ расподелом је $M_X(t) = E(e^{t^Tx}) = e^{\mu^Tt + \frac{1}{2}t^T\Sigma t}, \ \ t \in R$;
- X се може представити у облику $X = AZ + \mu$ где је $AA^T = \Sigma$ а Z има стандардну вишедимензионалну нормалну расподелу.
- Нека $X = (X_1, ..., X_n)^T$ има вишедимензиону нормалну расподелу. Тада су компоненте вектора независне акко су некорелисане (коваријациона матрица је дијагонална матрица).
- X има вишедимензиону нормалну расподелу акко за сваки вектор a (различит од нуле) a^TX има једнодимензионалну нормалну расподелу.

0.3.2 χ^2 расподела

Дефиниција 0.3.2. *Нека су* $X_1, ..., X_k$ *независне случајне величине са* $N(\theta_1, 1)$,...,$N(\theta_k, 1)$ *расподелма ,редом. Тада случајна величина*

$$Y = \sum_{j=1}^{k} X_j^2 \tag{3}$$

има $\chi_k^2(\mu)$ *расподелу, где је параметар положаја* $\mu = \sum_{j=1}^{k} \theta_j^2$. *Уколико је* $\mu = 0$ *параметар положаја ћемо изоставити у нотацији.*

Теорема 0.3.1. *Случајна величина* Y *дефинисана са* (3) *се може представити у облику збира две независне случајне величине од којих једна има* $\chi_1^2(\mu)$ *а друга* χ_{k-1}^2 *расподелу.*

Приметимо да заправо ова теорема оправдава дефиницију χ^2 расподеле јер сугерише да расподела од Y зависи само од степени слободе и μ.

Доказ. Нека је $B = [b_{ij}]$ ортогонала матрица тако да је $b_{1j} = \theta_j \mu^{-\frac{1}{2}}$, за $j = 1, 2..., k$. Нека је $W = BX$. Тада W има $N(B\theta, BIB^T)$, односно $N(B\theta, I)$ јер је B ортогонална матрица. Одавде је јасно да су компоненте вектора W међусобно независне. Приметимо да је $EW_1 = \sum_{j=1}^{k} b_{1j}\theta_j = \mu^{\frac{1}{2}}$, и да је $EW_i = \sum_{j=1}^{k} b_{ij}\theta_j = 0$. . Друга једнакост важи пошто је матрица B ортогонална.

$$Y = X^T X = W_1^2 + \sum_{i=2}^{k} W_i^2$$

Из ове репрезентације јасно следи тврђење теореме. □

0.3.3 Фишерова расподела

Нека $X \sim \chi_{n_1}^2(\mu)$ и $Y \sim \chi_{n_2}^2$ и независне су. Тада

$$\frac{\frac{X}{n_1}}{\frac{Y}{n_2}}$$

има Фишерову $F_{n_1,n_2}(\mu)$ расподелу.

0.3.4 Студентова расподела

Нека X има нормалну $N(\theta, 1)$ расподелу и Y има χ^2_m расподелу и независне су. Тада

$$\frac{X}{\sqrt{\frac{Y}{m}}}$$

има Студентову $t_m(\theta)$ расподелу, где је θ параметар положаја.

0.4 Расподела квадратне форме

Теорема 0.4.1 (Кохран). *Нека су $X_1, ..., X_n$ независне $\mathcal{N}(0, \sigma^2)$ случајне величине и нека је*

$$\sum_{i=1}^{n} X_i^2 = \sum_{j=1}^{k} Q_j,$$

где је Q_j квадратна форма дефинисана са $Q_j = X^T A_j X$, за $j = 1, 2, ..., k$, при чему је $R(A_j) = r_j$. Тада је $\sum_{j=1}^{k} r_j = n$ ако и само ако

1. *$Q_1, ..., Q_k$ су независне случајне величине и*
2. *Q_j/σ^2 има $\chi^2_{r_j}$ расподелу.*

Може се формулисати и општије тврђење:

Теорема 0.4.2 (Кохран). *Нека случајни вектор X има $\mathcal{N}(\theta, \sigma^2 I)$ расподелу и нека је*

$$\sum_{i=1}^{n} X_i^2 = \sum_{j=1}^{k} Q_j,$$

где је Q_j квадратна форма дефинисана са $Q_j = X^T A_j X$, за $j = 1, 2, ..., k$, при чему је $R(A_j) = r_j$. Тада је $\sum_{j=1}^{k} r_j = n$ и $\sum_{j=1}^{k} \mu_j = \theta^T \theta$ ако и само ако

1. *$Q_1, ..., Q_k$ су независне случајне величине и*
2. *Q_j/σ^2 има $\chi^2_{r_j}(\mu_j)$ расподелу, при чему је $\mu_j = \theta^T A_j \theta$.*

Доказаћемо само прву теорему. Друга се доказује аналогно. Пре него што се упустиомо у доказ наводимо једну њену опште познату последицу.

Последица 0.4.1. *Узорачка средина $\bar{X}$ и поправљена узорачка дисперзија S^2 су незавине случајне величине, и $(n-1)S^2/\sigma^2$ има χ^2_{n-1} расподелу.*

Без умањења општости можемо претпоставити да је $\sigma^2 = 1$.

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2.$$

Одавде је

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X})^2 = X^T\Big(I - \frac{J}{n}\Big)X + X^T\frac{J}{n}X,$$

где је J $n \times n$ матрица чији су сви елементи јединице. Приметимо да лева страна једнакости има χ_n^2 расподелу. Даље, $R(I - \frac{J}{n}) = n-1$ јер је, с једне стране $R(I - \frac{J}{n}) \geq R(I) - R(\frac{J}{n})$, а с друге, из једнакости $(I - \frac{J}{n})1 = 0$ закључујемо да је $R(I - \frac{J}{n}) \leq n-1$. Како је $R(\frac{J}{n}) = 1$ следи тврђење.

У доказу Кохранове теореме ћемо користити следећу лему.

Лема 0.4.1. *Нека су $x_1, ...x_n$ реални бројеви. Претпоставимо да се сума $\sum_{i=1}^{n} x_i^2$ може представити као збир k квадратних форми $\sum_{j=1}^{k} Q_j$ где је $Q_i = x^T A_i x$ и $R(A_i) = r_i$ за $i = 1, ..., k$. Ако је $\sum_{i=1}^{k} r_i = n$ тада постоји ортогонална матрица M, таква да за $x = My$ важи*

$$\begin{aligned}
Q_1 &= y_1^2 + \cdots y_{r_1}^2\\
Q_2 &= y_{r_1+1}^2 + \cdots y_{r_1+r_2}^2\\
&\cdots\\
Q_k &= y_{n-r_k+1}^2 + \cdots y_n^2
\end{aligned}$$

Доказ. Довољно је показати лему за $k = 2$. Тада је

$$Q = x^T x = x^T A_1 x + x^T A_2 x.$$

Постоји ортогонална матрица M таква да је $M^T A_1 M = D_1$ где је D_1 дијагонална матрица. Без умањења општости можемо претпоставити да су од сосвене вредности поређане тако да су $\lambda_1, ...\lambda_{r_1}$ различите од нуле а остале нула. Нека је $x = My$. Тада је

$$x^T x = y^T M^T M y = y^T y.$$

Даље је

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{r_1} \lambda_i y_i^2 + y^T M^T A_2 M y.$$

Одавде је

$$\sum_{i=1}^{r_1}(1-\lambda_i)y_i^2 + \sum_{i=r_1+1}^{n} y_i^2 = y^T M^T A_2 M y.$$

Како је $R(A_2) = n - r_1$ закључујемо да је $\lambda_1 = \cdots = \lambda_{r_1} = 1$ одакле следи тврђење. $\square$

Важно је приметити да све квадратне форме које учествују у репрезентацији садрже различите y_i-ове. Независност Q_1,..., Q_k у тврдјењу Кохранове теореме је последица овога. Да бисмо доказали Кохранову теорему потребно је још да приметимо да када применимо ортогоналну трансформацију на случајан вектор са нормалном расподелом са независним компонентама добијамо опет случајан вектор са нормалном расподелом са независним компонентама. Одавде се добијају расподеле одговарајућих квадратних форми.

Следећеу теорему наводимо без доказа.

Теорема 0.4.3. *Нека случајни вектор X има $N(\theta, \sigma^2 I)$ расподелу и нека је $Q_1 = X^T A_1 X$ и $Q_2 = X^T A_2 X$, где су A_1 и A_2 две симетричне матрице. Тада су Q_1 и Q_2 независне ако и само ако је $A_1 A_2 = 0$.*

0.1. Нека су $X_1, ..., X_n$ независне и једнако расподељене случајне величине тако да је $EX_1 = 0$ и $DX_1 = \sigma^2 < \infty$. Нека је $Y_i = X_i - \bar{X}$, за $i = 1, 2, ..., n$. Наћи коваријациону матрицу случајног вектора $Y = (Y_1, ..., Y_n)^T$.

0.2. Нека су $X_1, ..., X_{n_1}$ независне са $N(\mu_1, \sigma_1^2)$ и $Y_1, ..., Y_{n_2}$ независне са $N(\mu_2, \sigma_2^2)$ расподелом. Наћи расподелу селедећих статистика:

а) $\frac{(\bar{X}-\bar{Y}-\delta)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$, где је δ произвољна константа.

б) $\frac{n_1(\bar{X}-\mu_1)^2}{\sigma_1^2} + \frac{n_2(\bar{Y}-\mu_2)^2}{\sigma_2^2}$

0.3. Нека су $X_1, ..., X_{n_1}$ независне са $N(\mu_1, \sigma^2)$ и $Y_1, ..., Y_{n_2}$ независне са $N(\mu_2, \sigma^2)$ расподелом. Наћи расподелу селедећих статистика:

а) $\frac{S_1^2(n_1-1)+S_1^2(n_2-1)}{\sigma^2}$

0.4. Нека је $(X_1, Y_1), ..., (X_n, Y_n)$ прост случајан узорак из дводимензионалне нормалне расподеле са параметрима $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Одредити константу C тако да статистика

$$T = C\frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\sum_{i=1}^{n}(X_i - Y_i - \bar{X} + \bar{Y})^2}}$$

има Студентову $t_m(\theta)$. Изразити m и θ у функцији од параметара расподеле и константе δ.

0.5. Доказати теорему 0.4.3.

Поглавље 1

Линеарни модели

1.1 Проста линеарна регресија

У уводном поглављу смо претпостављали да имамо један предиктор. Такав модел се назива *прости линеарни регресиони модел.* Видели смо нека лепа својства која модел поседује при чему нисмо наводили које претпоставке модел треба да зодовољава да бисмо га уопште разматрали.

Нека је

$$Y_i = ax_i + b + \varepsilon_i, \quad i = 1, 2, ..., n$$

прост линеарни модел, при чему шум задовољава следеће услове Гаус-Маркова

1. центрираност $E\varepsilon_i = 0, \quad i = 1, 2, ..., n$
2. некорелисаност $E\varepsilon_i\varepsilon_j = 0, \ i \neq j$;
3. хомоскедастичност $D\varepsilon_i = \sigma^2 > 0$;
4. x_i и ε_j су независни за свако i, j.

Тада за оцене добијене методом најмањих квадрата

$$\begin{aligned}
\hat{a} &= \frac{\sum_{i=1}^n Y_i x_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{nS_x^2} \\
\hat{b} &= \bar{y} - \hat{a}\bar{x} = \frac{\sum_{i=1}^n Y_i(S_x^2 - \bar{x}(x_i - \bar{x}))}{nS_x^2},
\end{aligned}$$

важе следећа својства:

1. $E\hat{a} = 0$ и $E\hat{b} = 0$;

2. $$D(\hat{a}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$D(\hat{b}) = \frac{\sigma^2}{n}\Big(1 + \frac{\bar{x}}{S_x^2}\Big);$$

Дакле, оцене параметара модела су непристрасне и постојане. Видимо да у изразима за дисперзије вигурише непознати параметар σ^2. Зато морамо и за њега наћи одговарајућу оцену.

$$\begin{aligned}
E(SSE) &= E(SST) - E(SSR)\\
E(SST) &= E(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2)\\
&= \sum_{i=1}^{n}(\sigma^2 + (aX_i + b)^2) - \frac{1}{n}(\sum_{i=1}^{n} EY_i^2 + 2\sum_{1\le i<j<n} EY_iEY_j)\\
&= (n\sigma^2 + \sum_{i=1}^{n}((aX_i + b)^2))\frac{(n-1)}{n} - \frac{2}{n}(\sum_{1\le i<j\le n}(aX_i + b)(aX_j + b))\\
&= \sigma^2(n-1) + nS_x^2a^2\\
E(SSR) &= \sum_{i=1}^{n}(x_i - \bar{x})^2E\hat{a}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2(a^2 + \frac{\sigma^2}{nS_x^2})\\
&= nS_x^2(a^2 + \frac{\sigma^2}{nS_x^2})\\
E(SSE) &= \sigma^2(n-2)
\end{aligned}$$

Одавде добијамо да је непристрасна оцена за σ^2

$$\hat{\sigma}^2 = \frac{SSE}{n-2}.$$

Уколико се у модел уведе додатна претпоставка да је шум Гаусов, односно да ε_i има нормалну $\mathcal{N}(0, \sigma^2)$, добијене оцене имају многа друга лепа својства. Прво, приметимо да су оцене за a и b линеарне комбинације независних случајних величина са нормалним расподелама па и саме оцене имају нормалне $\mathcal{N}(a, D\hat{a})$ и $\mathcal{N}(b, D\hat{b})$ расподеле. У наставку курса ћемо показати да $\hat{a}$ и $\hat{b}$ су независне од $\hat{\sigma}^2$ па закључујемо да

$$\frac{\hat{a} - a}{\frac{\hat{\sigma}}{\sqrt{nS_x^2}}} \sim t_{n-2}$$

$$\frac{\hat{b} - b}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nS_x^2}}} \sim t_{n-2}$$

Сада се могу правити интервали поверења за a и b и тестирати хипотезе у вези са њиховим параметрима. Приметимо да $H_0 : a = 0$ заправо значи да утицај предиктора није значајан.

Прогнозирана вредност зависне променљиве Y_0 и средње вредности зависне променљиве EY_0 у тачки x_0 је

$$\hat{Y}_0 = \hat{a}x_i + \hat{b}.$$

Користећи исте аргументе као до сада, можемо показати да

$$\frac{\hat{Y}_0 - EY_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{nS_x^2}}} \sim t_{n-2}$$

$$\frac{\hat{Y}_0 - Y_0}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{nS_x^2}}} \sim t_{n-2}.$$

Сада можемо правити интервале предвиђања.

1.2 Вишеструка линеарна регресија

Међутим, неретко се дешава да треба узети у обзир неколико предиктора.

Претпоставимо да имамо p предиктора $x_1, x_2, ..., x_p$. Тада се линеарни модел може записати у облику

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1.1}$$

или у облику

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{1.2}$$

при чему је n обим узорка који имамо на располагању.

Најчешће се овај модел записује у векторском облику.

1.3 Матрични запис линеарног модела

Модел (1.1) се може записати у облику

$$Y = X\beta + \varepsilon,$$

где су

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix},$$

X се назива предиктор (независна промењлива), а Y регресанд (зависна промењлива). У случају модела (1.2) је

$$X = \begin{pmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_n^T \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},$$

Матрица X се назива *дизајн матрица.*

Случајност модела потиче од случајних грешки $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ за које се претпоставља центрираност, хомоскедастичност и некорелисаност. Дакле претпостављамо да важи:

1. $E(\varepsilon) = 0$;
2. $D(\varepsilon) = \sigma^2 I$;
3. X и ε су независни случајни вектори.

Пример 1.3.1. *У случају просте линеарне регресије је*

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \beta = \begin{pmatrix} b \\ a \end{pmatrix}.$$

Пример 1.3.2. *Желимо да видимо какав је утицај неког фактора на средњу вредност мереног обележја* Y. *За почетак, претпоставимо да имамо само један фактор који има две категорије. Нека је* $x_i = 1$ *уколико меримо вредност обелжја* Y *на узорку из прве категорије. У супротном нека је* $x_i = 0$. *Даље, претпоставимо да имамо тачно* k *елемената узорка из прве категорије. Тада модел можемо приказати у облику.*

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Примитимо да је ово заправо ANOVA модел.

Пример 1.3.3. *Понекад, на основу графичког приказа података, имамо разлога да верујемо да би било боље, уместо линеарне зависности, претпоставити квадратну зависност.*

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, 2, ..., n.$$

Овај модел можемо написати на сличан начин у векторском облику. Сада је

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \\ 1 & x_n & x_n^2 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

Најприроднији начин да се оцене коефицијенти модела β је метод најмањих квадрата, односно

$$\hat{\beta} := \arg \min_{b \in R^{p+1}} \|Y - Xb\|^2.$$

Уколико је $R(X) = p + 1$ онда је $X^T X$ инвертибилна и добија се да је

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Оцењена вредност је тада

$$\hat{Y} = X(X^T X)^{-1} X^T Y = HY$$

.

Матрица H се назива hat matrix. Приметимо да је матрица H пројектор, тако да $\hat{Y}$ представља ортогоналну пројекцију вектора Y на раван генерисану са X. Резидуали модела се могу приказати у облику

$$e = Y - \hat{Y} = (I - H)Y.$$

Одавде је

$$\begin{aligned} E(e) &= (I - H)E(Y) = 0 \\ Cov(e) &= \sigma^2(I - H)(I - H)^T = \sigma^2(I - H) \end{aligned}$$

Посматрајмо суму квадрата одступања од модела

$$SSE = \sum_{i=1}^{n} e_i^2 = e^T e = Y^T(I - H)Y = Y^T Y - Y^T HY.$$

Означимо са $M = (I - H) = [m_{ij}]$. Њено очекивање је

$$\begin{aligned} E(SSE) = E(\sum_{ij} m_{ij} Y_i Y_j) &= \sum_{ij} m_{ij} E(Y_i Y_j) = \sum_{ij} m_{ij} E(\varepsilon_i \varepsilon_j) \\ &= \sum_{i} m_{ii} E(\varepsilon_i^2) = \sigma^2 \sum_{i} m_{ii} \\ &= \sigma^2 tr(M) = \sigma^2(tr(I) - tr(H)) = \sigma^2(n - tr((X^T X)(X^T X)^{-1})) \\ &= \sigma^2(n - tr(I_{p+1})) = \sigma^2(n - p - 1). \end{aligned}$$

Одавде закључујемо да је непристрасна оцена за σ^2 управо $(\sum_{i=1}^{n} e_i^2)/(n - p - 1)$.

Резидуали и оцењене вредности су некорелисани $\sum(Y_i - \hat{Y})(\hat{Y} - \bar{Y}) = 0$ па је

$$\begin{aligned} \sum_{i=1}^{n}(Y_i - \bar{Y})^2 =& \sum_{i=1}^{n}(Y_i - \hat{Y})^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\ SSTO =& SSE + SSR \end{aligned} \tag{1.3}$$

SSE је необјашњено одступање (потиче од модела), SSR објашњено одступање а $SSTO$ укупно одступање. У матричном облику се одступања могу представити на следећи начин:

$$\begin{aligned} SSTO &= Y^T(I - \frac{J}{n})Y, \\ SSE &= Y^T(I - H)Y \\ SSR &= Y^T(H - \frac{J}{n})Y, \end{aligned}$$

где је матрица J $n \times n$ матрица са свим јединицама. Приликом доказа разлагања (1.3) коришћена је следећа лема.

Лема 1.3.1. *Нека су $\{e_i\}$ резидуали линеарног модела. Тада је*

$$\sum_{i=1}^{n} e_i = 0.$$

Прооф. Претпоставимо да је $\sum_{i=1}^{n} e_i = cn \neq 0$. Тада је

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(e_i - c + c)^2 = \sum_{i=1}^{n}(e_i - c)^2 + nc^2 + 2c\sum_{i=1}^{n}(e_i - c) \geq \sum_{i=1}^{n}(e_i - c)^2$$
$$= \sum_{i=1}^{n}(Y_i - (1, x_i^T)\hat{\beta} - c)^2.$$

Одавде закључујемо да вектор $\hat{\beta}$ не минимизира суму квадрата одступања што је супротно претпоставци, па c мора бити 0.

Из методолошких разлога приказаћемо још један начин да се ово покаже. Приметимо да су резидуали ортогонални на простор генерисан са X.

$$e^T X = Y^T(I - H)X = Y^T(X - H^T X) = \mathbf{0}$$

Како је први ред матрице X^T вектор јединица, следи тврђење леме. □

Из формуле (1.3) је природно да о квалитету модела говоримо на основу количине објашњених одступања моделом односно да посматрамо *коефицијент детерминације* $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$. Видимо да је $0 \leq R^2 \leq 1$. У пракси је потребно да је R^2 бар 0.62. Са $R = \sqrt{R^2}$ је дефинисан вишеструки коефицијент корелације. Приметимо да кад имамо само један предиктор онда је $R = |\rho_{xy}|$.

Даље, важи да је $R(I - \frac{J}{n}) = n - 1$, као и $R(I - H) = n - p - 1$ јер је

$$R(I - H) = tr(I - H) = n - p - 1.$$

С обзиром на то, да коефицијент детерминације увек расте са порастом броја предиктора понекад се уместо њега користи његова модификација која узима у обзир непристрасне оцене одговарајућих грешака,дефинисана са

$$R_A^2 = 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SSTO}{n-1}}.$$

У пракси је примећено да у случају малих узорака вредности R^2 могу бити велике чак и кад модел није довољно квалитета. За разлику од R^2, R_A^2 може бити негативан.

$R(X)$ не мора бити једнак баш $p+1$. На пример меримо температуру у цезијусима и фарехајтима, или бележимо број поена на предиспитним обавезама, на испиту, и укупан број поена. Може се десити и да је број променљивих већи или једнак од броја обзервација. Ако је $p+1=n$ ради се о *сатурираном* моделу, док ако је $p>n-1$ кажемо да је модеј *суперсатуриран*. У овом случају, кеофицијенти модела се не могу јединствено одредити.

Најбоље је обратити пажњу у прелиминарној анализи и уклонити променљиве које ису неопходне.

Ако је X ортогонална матрица онда се оцена за β поклапа са вектором који би се добио када бисмо посматрали $p+1$ одговарајућих простих линеарних регресија.

Вратимо се сада особинама оцене коефицијената модела методом најмањих квадрата.

Последица 1.3.1. *Уколико $tr((X^TX)^{-1}) \to 0$, кад $n \to \infty$ $\hat{\beta}$ је постојана оцена параметра β.*

Важна особина оцене методом најмањих квадрата је садржана у следећој теореми.

Теорема 1.3.1. *Нека је $\hat{\beta}$ оцена методом најмањих квадрата и $\tilde{\beta} = AY$ нека друга линеарна непристрасна оцена за β. Тада је матрица $Cov(\tilde{\beta}|X) - Cov(\hat{\beta}|X)$ позитивно дефинитна.*

Прооф.

$$E(\hat{\beta}|X) = (X^TX)^{-1}X^TE(Y|X) = (X^TX)^{-1}X^TX\beta = \beta$$
$$Cov(\hat{\beta}|X) = (X^TX)^{-1}X^T\sigma^2IX(X^TX)^{-1} = (X^TX)^{-1}\sigma^2$$

Нека је $B=(X^TX)^{-1}X^T$ и нека је $A=B+C$. Из услова непристрасности је $E(\tilde{\beta}|X) = AX\beta = \beta$. Односно $I = AX = (B+C)X$. Како је $BX=I$ закључујемо да је $CX=0$.
Даље је

$$Cov(\tilde{\beta}|X) = AD(Y|X)A^T = (B+C)(B^T+C^T)\sigma^2 = Cov(\hat{\beta}|X) + BC^T + CB^T + CC^T.$$

Како је

$$CB^T = CX(X^TX)^{-1} = 0$$
$$BC^T = (X^TX)^{-1}X^TC^T = (X^TX)^{-1}(CX)^T = 0$$

следи тврђење. □

Претпоставимо да је циљ да оценимо неку линеарну комбинацију параметара модела или више таквих линеарних комбинација, односно да оценимо $l^T\beta$ у првом случају, односно $(L^T\beta)$ у другом случају. Задржимо се за тренутак на првом случају. Други је аналоган. Претпоставимо да постоји непристрасна линеарна оцена (cY) оцена за $l^T\beta$. Тада се од свих таквих оцена оцена са најмањом дисперзијом назива најбоља линеарна непристрасна оцена и означава са BLUE (Best Linear unbiased estimate).

Слично као претходна теорема се показују и следеће теореме. Прва говори о оцени линеарне функције параметара модела $(l^T\beta)$ а друга о оцени више линеарних функција параметара модела $(L^T\beta)$.

Теорема 1.3.2 (Гаус-Марков). *$l^T\hat{\beta}$ је $BLUE$ за $l^T\beta$.*

Теорема 1.3.3 (Гаус-Марков). *$L^T\hat{\beta}$ је $BLUE$ за $L^T\beta$.*

1.4 Статистичко закључивање у Гаусовом моделу

У овом поглављу претпоставићемо да грешке модела представљају Гаусов шум, односно да $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$.

1.4.1 Тестирање линеарне хипотезе

Желимо да тестирамо нулту хипотезу $H_0 : C\beta = \gamma$, где је C $m \times (p+1)$ матрица. На пример, уколико желимо да тестирамо нулту хопотезу да нема утицаја слободног члана узећемо да је $C = (1, 0, ..., 0)$ и $\gamma = 0$. Или, уколико желимо да тестирамо нулту хипотезу да ниједан предиктор нема утицаја на Y, $C = (0, I_p)$ и $\gamma = 0$.

За тестирање наше нулте хипотезе користићемо тест количника веродостојности.

1.4.2 Метод максималне веродостојности

Претпоставимо да ε има $N(0, \sigma^2 I_n)$ расподелу. Тада је функција веродостојности

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{(Y-\beta X)^T(Y-\beta X)}{2\sigma^2}}.$$

Одавде видимо да се оцена методом максималне веродостојности за β поклапа са оценом методом најмањих квадрата, односно $\hat{\beta} = (X^TX)^{-1}X^TY$. Слично, оцена за σ^2 је

$$\hat{\sigma}^2 = \frac{1}{n}(Y - \hat{\beta}X)^T(Y - \hat{\beta}X)$$

У претходном одељку видели смо да ова оцена није непристрасна оцена за σ^2.

Максимална вредност функције веродостојности је тада

$$L(\hat{\beta}, \hat{\sigma^2}) = \frac{1}{(2\pi\hat{\sigma}^2)^{\frac{n}{2}}} e^{-\frac{1}{2n}} \sim \hat{\sigma}^{-n}$$

Како је $\hat{\beta}$ линеарна трансформација случајног вектора са нормалном расподелом, закључујемо да

$$\hat{\beta} \sim N(\beta, (X^TX)^{-1}\sigma^2).$$

Да бисмо применили тест количника веродостојности неопходно је да нађемо оцену параметара уз услов $C\beta = \gamma$. Слично као у поставци проблема малопре, ово је еквивалентно са тражењем минимума $S(\beta) = \frac{(Y-X\beta)^T(Y-X\beta)}{n}$ уз услов $C\beta - \gamma = 0$ (зато што у услову не фигурише σ^2). Лагранжова функција за решавање овог екстремалног проблема је

$$\mathcal{L}(\beta, a) = (Y - \beta X)^T(Y - \beta X) - a^T(C\beta - \gamma).$$

Систем нормлних једначина чије решење тражимо је

$$\begin{aligned} \frac{\partial}{\partial\beta}\mathcal{L}(\beta, a) &= -2X^TY + 2X^TX\beta - C^Ta = 0 \\ \frac{\partial}{\partial a}\mathcal{L}(\beta, a) &= \gamma - C\beta = 0. \end{aligned}$$

Добија се да је

$$\hat{\beta}_0 = \hat{\beta} + (XX^T)^{-1}C^T[C(XX^T)^{-1}C^T]^{-1}(\gamma - C\hat{\beta})$$

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{(Y - X\hat{\beta}_0)^T(Y - X\hat{\beta}_0)}{n}\\ &= \frac{1}{n}\Big((Y - X\hat{\beta})^T(Y - X\hat{\beta}) + (C\hat{\beta} - \gamma)^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta} - \gamma)\Big)\end{aligned}$$

Максимална вредност функције веродостојности у овом случају је

$$L(\hat{\beta}_0, \hat{\sigma^2}_0) \sim \hat{\sigma}_0^{-n}$$

Претпоставимо да је $H_1 : \beta X \neq \gamma$. Тада је количник веродостојности

$$\begin{aligned}\lambda &= \frac{\max\limits_{H_1} L(\beta, \sigma^2)}{\max\limits_{H_0} L(\beta, \sigma^2)} = \frac{L(\hat{\beta}, \hat{\sigma}^2)}{L(\hat{\beta}_0, \hat{\sigma}_0^2)} = \Big(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\Big)^{\frac{n}{2}}\\ &= \Big(1 + \frac{(C\hat{\beta} - \gamma)^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta} - \gamma)}{e^Te}\Big)^{\frac{n}{2}}\end{aligned}$$

Критична област за тестирање је $W = \{\lambda > c\}$.

Претпоставка о нормалном моделу је јако важна за коришћење теста количника веродстојности јер се тест статистика може приказати у погодном облику као количник две независне статистике са χ^2 расподелама, односно у функцији од статистике која, под нултом хипотезом, има Фишерову расподелу.

Јасно је да $\hat{\beta} = (X^TX)^{-1}XY$ има нормалну $N(\beta, \sigma^2(X^TX)^{-1})$. На основу тога имамо да

$$C\hat{\beta} - \gamma \sim N_m(C\beta - \gamma, \sigma^2C(X^TX)^{-1}C^T).$$

Одавде, уколико важи H_0 закључујемо да

$$Q = (C\hat{\beta} - \gamma)^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta} - \gamma) \sim \sigma^2\chi_m^2.$$

Приметимо да се, уколико важи H_0, Q може приказати у облику

$$\begin{aligned}Q &= (C\hat{\beta} - C\beta)^T[C(X^TX)^{-1}C^T]^{-1}(C\hat{\beta} - C\beta)\\ &= (C(X^TX)^{-1}X^T\varepsilon)^T[C(X^TX)^{-1}C^T]^{-1}C(X^TX)^{-1}X^T\varepsilon\\ &= \varepsilon^TX(X^TX)^{-1}C^T[C(X^TX)^{-1}C^T]^{-1}C(X^TX)^{-1}X^T\varepsilon\\ &= \varepsilon^TP\varepsilon,\end{aligned}$$

где је $P = X(X^TX)^{-1}C^T[C(X^TX)^{-1}C^T]^{-1}C(X^TX)^{-1}X^T$.

Даље, применом Кохранове теореме добијамо да

$$e^Te = SSE = \varepsilon^TM\varepsilon \sim \sigma^2\chi^2_{n-p-1}$$

Да бисмо нашли расподелу тест статистике потребно је још да покажемо да је $PM = 0$ и применимо теорему 0.4.3. Из једнакости $X^TM = 0$ добијамо $PM = 0$ и закључујемо да су бројилац и имеилац у тест статистици независне случајне величине па

$$\frac{\frac{Q}{m}}{\frac{e^Te}{n-p-1}} \sim F_{m,n-p-1}. \tag{1.4}$$

Сада је могуће применити тест количника веродостојности.

Пример 1.4.1. *Претпоставимо да тестирамо нулту хипотезу* $H_0 : \beta_1 = \beta_2 = \cdots \beta_p = 0$ *против комплементарне хипотезе. Тада је*

$$L(\hat{\beta}_0, \hat{\sigma}_0^2) \sim \left(\frac{SSTO}{n}\right)^{-\frac{n}{2}}.$$

Одавде, је

$$\lambda = \left(\frac{SSTO}{SSE}\right)^{\frac{n}{2}} = \left(1 + \frac{SSR}{SSE}\right)^{\frac{n}{2}}.$$

Применом Кохранове теореме, или теореме (0.4.3) *добија се да*

$$\frac{\frac{SSR}{p}}{\frac{SSE}{n-p-1}} \sim F_{p,n-p-1}.$$

Пример 1.4.2. *У случају да желимо да тестирамо хипотезу да је* $\beta_k = 0$ *већ смо видели како да одаберемо матрицу* C. *Тада*

$$\frac{Q}{\frac{e^Te^T}{n-p-1}} \sim F_{1,n-p-1}.$$

У овом посебном случају овај тест са Студентовом тест статистиком коју ћемо представити у наредном поглављу.

Литература

[1] J. J. Faraway. *Linear models with R*. CRC press, 2014.

[2] Roger Koenker. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, 17(1):107–112, 1981.

[3] J. Neter, M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. *Applied linear statistical models*, volume 4. Irwin Chicago, 1996.

[4] A. Sen and M. Srivastava. *Regression analysis: theory, methods, and applications*. Springer Science & Business Media, 2012.