evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERZITA OBRANY

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Regresní modely

1 Regresní přímka

Princip regresní analýzy nejdříve vysvětlíme na jednoduchém modelu dvou náhodných veličin X a Y, kde Y bude vysvětlovaná proměnná a X bude vysvětlující proměnná (**regresor**). Budeme předpokládat, že mezi vysvětlovanou proměnnou Y a vysvětlující proměnnou X platí přibližně lineární vztah. Měření nebo pozorování veličiny Y může být zatíženo náhodnou chybou e.

$$Y = \beta_1 + \beta_2 X + e,$$

kde $\beta_1,\ \beta_2$ jsou neznámé parametry (neznámé reálné konstanty), Y a e jsou náhodné veličiny a X je daná reálná proměnná. Dále předpokládáme, že při hodnotách $x_1, x_2, \ldots, x_n$ proměnné X pozorujeme hodnoty $y_1, \ldots, y_n$ proměnné Y zatížené chybami $e_1, \ldots, e_n$. Pozorování vyhovují modelu

$$y_i = \beta_1 + \beta_2 x_i + e_i,\ \ i = 1, \ldots,\ n.$$

O chybách $e_1, \ldots, e_n$ předpokládáme, že jsou to nezávislé náhodné veličiny, že jsou **nesystematické**, tj. střední hodnota $Ee_i = 0$, a **homogenní**, tj. že mají stejný rozptyl $De_i = \sigma^2,\ i = 1, \ldots, n$. Cílem je najít odhad parametrů $\beta_1,\ \beta_2$ a σ^2. Použijeme k tomu **metodu nejmenších čtverců**. Označíme

$$S^2(\beta_1, \beta_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (\beta_1 + \beta_2 x_i))^2$$

součet čtverců náhodných chyb e_i a odhady $\widehat{\beta}_1, \widehat{\beta}_2$ parametrů β_1, β_2 stanovíme tak, aby součet čtverců chyb $S^2(\beta_1, \beta_2)$ nabyl minimální možné hodnoty. Z matematiky je známo, že nutnou podmínkou pro existenci extrému funkce dvou a více proměnných je nulovost prvních parciálních derivací, tj. v našem případě

$$\frac{\partial S^2(\beta_1, \beta_2)}{\partial \beta_1} = \frac{\partial S^2(\beta_1, \beta_2)}{\partial \beta_2} = 0,$$

podmínku postačující pro minimum nemusíme vyšetřovat, neboť funkce $S(\beta_1, \beta_2)$ je ryze konvexní. Dostáváme tedy

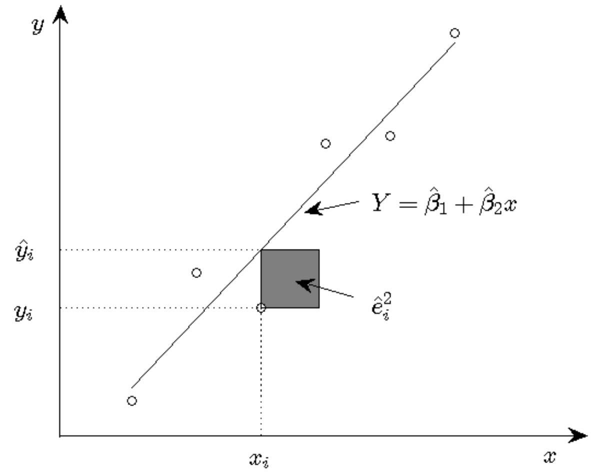
$$\frac{\partial S^2(\beta_1, \beta_2)}{\partial \beta_1} = 2\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)(-1) = 0,$$
$$\frac{\partial S^2(\beta_1, \beta_2)}{\partial \beta_2} = 2\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)(-x_i) = 0.$$

odkud získáme tzv. **soustavu normálních rovnic**

$$\beta_1 n + \beta_2 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i,$$
$$\beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

Operační program Vzdělávání pro konkurenceschopnost
Název projektu: Inovace magisterského studijního programu Fakulty ekonomiky a managementu
Registrační číslo projektu: CZ.1.07/2.2.00/28.0326

PROJEKT JE SPOLUFINANCOVÁN EVROPSKÝM SOCIÁLNÍM FONDEM A STÁTNÍM ROZPOČTEM ČESKÉ REPUBLIKY.



Obr. 1: Lineární regresní model – přímka

Vyřešíme-li tuto soustavu (např. Cramerovým pravidlem), obdržíme odhady parametrů

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \ \widehat{\beta}_2 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}.$$

Tyto odhady lze také vyjádřit ve tvaru

$$\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2 \overline{x} = \overline{y} - \frac{s_{xy}}{s_x^2}\overline{x}, \ \ \widehat{\beta}_2 = \frac{s_{xy}}{s_x^2},$$

kde $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ a $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$ jsou výběrové průměry, $s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$ je výběrový rozptyl a $s_{xy} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$ je výběrová kovariance. Přímku o rovnici $\widehat{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 X$ nazýváme **regresní přímkou**, $\widehat{\beta}_1,\ \widehat{\beta}_2$ jsou tzv. **regresní koeficienty**. Vypočtené regresní koeficienty $\widehat{\beta}_1,\ \widehat{\beta}_2$ jsou nevychýlenými odhady neznámých parametrů $\beta_1,\ \beta_2$. Dále hodnota $\widehat{y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_i$ je predikovaná hodnota y v bodě x_i a veličiny $\widehat{e}_i = y_i - \widehat{y}_i = y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i$ nazýváme **rezidua**. Dále platí, že minimální hodnota součtu čtverců $S^2(\beta_1,\ \beta_2)$ je rovna

$$S_e = S^2\left(\widehat{\beta}_1,\ \widehat{\beta}_2\right) = \sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - \widehat{\beta}_1 \sum_{i=1}^n y_i - \widehat{\beta}_2 \sum_{i=1}^n x_i y_i.$$

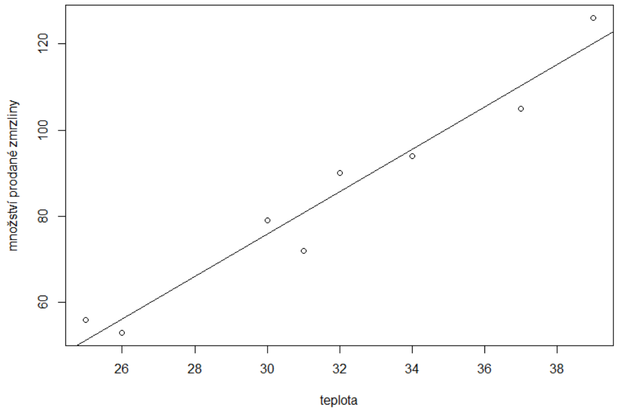
S_e nazýváme **reziduální součet čtverců**. Je možné ukázat, že veličina $s^2 = \frac{1}{n-2} S_e$ je nevychýleným odhadem rozptylu σ^2, a tedy platí $Es^2 = \sigma^2$.

Následující tabulka udává informaci o teplotě (ve stupních Celsia) v jednom městě a množství zmrzliny (v kilogramech) prodaných v osmi náhodně vybraných cukrárnách.

teplota	34	30	25	32	37	39	31	26
zmrzlina	94	79	56	90	105	126	72	53

Vysvětlovanou proměnnou je v tomto případě množství zmrzliny, vysvětlující proměnnou potom teplota ve městě. Metodou nejmenších čtverců odhadneme parametry regresní přímky

$$\widehat{y} = -71{,}789 + 4{,}918x.$$



Obr. 2: Regresní přímka – závislost množství prodané zmrzliny na teplotě

2 Lineární regresní model

Zobecníme předchozí výsledky a budeme předpokládat, že je potřeba modelovat nějakou sledovanou (hůře dostupnou či nesnadno měřitelnou) náhodnou veličinu Y (tzv. **vysvětlovaná veličina** nebo **odezva**) pomocí jiných snáze dostupných veličin $X_1, X_2, \ldots, X_k$ (nazývaných **vysvětlující proměnné** nebo **regresory**).

Vyjdeme ze situace, kdy příslušná statistická data obsahují n nezávislých pozorování vysvětlované proměnné Y a odpovídajících n pozorování každého z regresorů $X_1, X_2, \ldots, X_k$. Budeme předpokládat, že i-té pozorování vysvětlované proměnné Y lze modelovat rovnicí:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i, \tag{1}$$

kde

1. y_i je i-té pozorování Y, $i = 1, \ldots, n$,
2. x_{ij} je i-té pozorování regresoru X_j, $i = 1, \ldots, n$, $j = 1, \ldots, k$,
3. β_j, $j = 1, \ldots, k$, jsou neznámé parametry,
4. e_i, $i = 1, \ldots, n$, jsou neznámé náhodné chyby, které vznikají při pozorování vysvětlované proměnné Y a které nemůžeme přímo pozorovat ani měřit.

Přitom dále předpokládáme, že x_{ij} jsou pevně dané známé reálné hodnoty a veličiny Y_i a e_i jsou náhodného charakteru (náhodné veličiny). Na jejich pravděpodobnostní rozdělení klademe následující předpoklady:

(P1) Střední hodnota $Ee_i = 0$, $i = 1, \ldots,\ n$, tj. náhodné chyby jsou **nesystematické**.

(P2) Rozptyl $De_i = \sigma^2$, $i = 1, \ldots,\ n$, tj. náhodné chyby jsou **homogenní** se stejným neznámým rozptylem σ^2.

(P3) Kovariance $C(e_i, e_l) = 0$, $i \neq l$, $\ i,\ l = 1, \ldots,\ n$, tj. náhodné chyby jsou **nekorelované**.

Model daný rovnicí (1) spolu s předpoklady (P1), (P2), (P3) se nazývá **lineární regresní model** (LRM)[1]. Funkci, která popisuje závislost vysvětlované proměnné Y na regresorech $X_1, X_2, \ldots, X_k$ pak nazýváme **regresní funkcí**.

Odhad parametrů v lineárním regresním modelu (1) provedeme opět metodou nejmenších čtverců. Model nejdříve zapíšeme v maticovém tvaru. Označme:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

Pak model (1) lze vyjádřit jednoduchým zápisem

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$

Odhad neznámých parametrů pak stanovíme řešením soustavy lineárních rovnic

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}\text{– tzv. } \textbf{normální rovnice}.$$

Jejich řešení snadno nalezneme za předpokladu, že matice $\mathbf{X}'\mathbf{X}$ je regulární a tedy existuje inverzní matice $(\mathbf{X}'\mathbf{X})^{-1}$. Za tohoto předpokladu říkáme, že model je plné hodnosti. V modelu plné hodnosti lze řešení normálních rovnic zapsat ve tvaru

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}.$$

Pro reziduální součet čtverců zapsaný v maticovém tvaru pak dostaneme vyjádření

$$S_e = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}.$$

Dále budeme pracovat jenom s modely plné hodnosti. Uvedeme nyní dva příklady lineárních regresních modelů: regresní paraboly a modelu se dvěma lineárními regresory. Nejprve budeme uvažovat model, kdy vysvětlovaná proměnná Y je kvadratickou funkcí vysvětlující proměnné X, tvaru:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + e_i, \ \ i = 1, \ldots, \ n.$$

Zřejmě jde o speciální případ LRM (lineárního vzhledem k neznámým parametrům β_1, β_2, β_3). V maticovém zápisu tohoto modelu je:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}, \ \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix},$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{pmatrix}.$$

[1]Často se v lineárním regresním modelu předpokládá, že první regresor je konstanta, potom pozorované hodnoty $x_{i1} = 1, i = 1, \ldots, n$ a model má tvar $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i$.

2.1 Regresní parabola

Za předpokladu, že model je plné hodnosti, lze odhad $\widehat{\boldsymbol{\beta}}$ vektoru $\boldsymbol{\beta}$ získat řešením rovnic $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$ ve tvaru $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Potom lze reziduální součet čtverců S_e vyjádřit ve tvaru

$$S_e = \sum_{i=1}^{n} y_i - \widehat{\beta}_1 \sum_{i=1}^{n} y_i - \widehat{\beta}_2 \sum_{i=1}^{n} x_i y_i - \widehat{\beta}_3 \sum_{i=1}^{n} x_i^2 y_i$$

a odhad rozptylu σ^2 je $s^2 = S_e/(n-3)$.

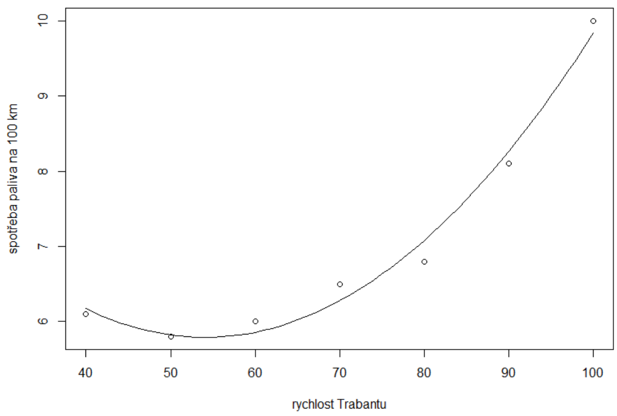
Příklad. U automobilu Trabant se měřila spotřeba paliva v litrech na 100 km (Y) v závislosti na jeho rychlosti (X).

Rychlost	40	50	60	70	80	90	100
Spotřeba	6,1	5,8	6,0	6,5	6,8	8,1	10,0

Odhadnutá parabolická regresní funkce má tvar

$$\widehat{y} = 11{,}39386 - 0{,}20726x + 0{,}001917x^2.$$

Předpokládejme, že vysvětlovaná proměnná Y může záviset na dvou regresorech X a Z (používáme označení



Obr. 3: Regresní parabola – závislost spotřeby paliva na rychlosti

X místo X_1 a Z místo X_2, které je v aplikacích tohoto typu časté). K dispozici je n nezávislých pozorování veličiny Y při daných n hodnotách veličin X a Z. Vyjdeme z modelu

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + e_i, \;\; i = 1, \ldots, n,$$

který je speciálním případem obecného lineárního regresního modelu $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.

2.2 Dva lineární regresory

Matice v modelu mají tvar

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{pmatrix}, \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n z_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i z_i \\ \sum_{i=1}^n z_i & \sum_{i=1}^n x_i z_i & \sum_{i=1}^n z_i^2 \end{pmatrix},$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n z_i y_i \end{pmatrix}.$$

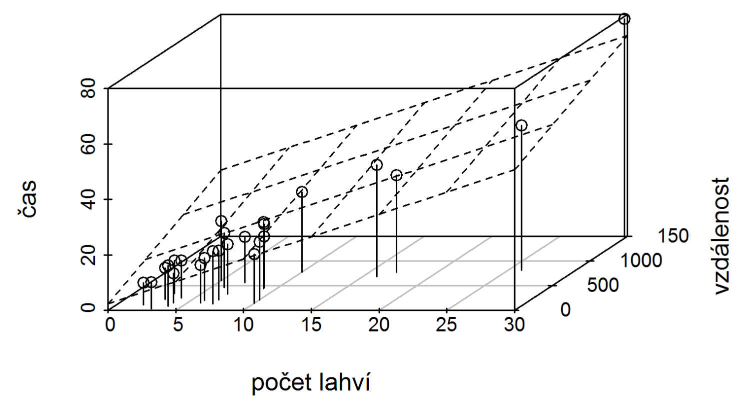
Pak užitím metody nejmenších čtverců dostaneme odhad $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

Příklad. Výrobce nealkoholických nápojů má zájem analyzovat potřebný čas k servisu (doplnění lahví případně malý servis zařízení) automatů na výdej lahví s těmito nápoji. Celkovou dobu doplnění lahví je třeba predikovat pomocí dvou dostupných proměnných: počet lahví, které je třeba doplnit do automatu, a vzdálenost, kterou musí údržbář ujít. Vysvětlovanou proměnnou je v tomto případě celkový čas, vysvětlující proměnné jsou počet doplněných lahví a vzdálenost.

čas	16,68	11,5	12,03	14,88	13,75	18,11	8	17,83	79,24	21,5
počet lahví	7	3	3	4	6	7	2	7	30	5
vzdálenost	560	220	340	80	150	330	110	210	1460	605
čas	40,33	21	13,5	19,75	24	29	15,35	19	9,5	35,1
počet lahví	16	10	4	6	9	10	6	7	3	17
vzdálenost	688	215	255	462	448	776	200	132	36	770
čas	17,9	52,32	18,75	19,83	10,75					
počet lahví	10	26	9	8	4					
vzdálenost	140	810	450	635	150					

Metodou nejmenších čtverců získáme odhad regresní funkce

$$\widehat{y} = 2,341 + 1{,}616x + 0{,}014z.$$



Obr. 4: Regrese se dvěma lineárními regresory – závislost času potřebného na servis na počtu případů doplňování automatu a vzdálenosti, kterou musí údržbář ujít

Některé typy lineárních regresních funkcí:

- přímková regrese $Y = \beta_1 + \beta_2 X$,

- hyperbolická regrese $Y = \beta_1 + \frac{\beta_2}{X}$,
- logaritmická regrese $Y = \beta_1 + \beta_2 \ln X$,
- parabolická regrese $Y = \beta_1 + \beta_2 X + \beta_3 X^2$
- polynomická regrese $Y = \beta_1 + \beta_2 X + \cdots + \beta_p X^p$

Některé typy nelineárních regresních funkcí:

- exponenciální regrese $Y = \beta_1 \beta_2^X$,
- mocninná regrese $Y = \beta_1 X^{\beta_2}$.

Příklady k procvičení

1. Byla zjištěna výška otců a výška jejich nejstarších synů [v cm].

otec	165	178	158	170	180	160	170	167	185	165	173	175
syn	162	184	163	170	189	165	177	170	187	176	171	183

   a) Sestrojte bodový graf.
   b) Určete regresní přímku a nakreslete její graf.
   c) Odhadněte průměrnou výšku syna při výšce otce 178 cm.

[Datový soubor: vyska_otec_syn.txt]

2. O 7 vybraných strojích v určitém podniku máme informace o jejich stáří (v letech) a týdenních nákladech na jejich údržbu (v Kč):

stáří stroje	1	1	3	3	5	6	7
náklady	35	52	81	105	100	125	120

   a) Sestrojte bodový graf.
   b) Určete regresní přímku a nakreslete její graf.
   c) Určete regresní logaritmickou křivku a nakreslete její graf.

[Datový soubor: stari_stroje_naklady.txt]

3. Zajímáme se o brzdnou dráhu 63 automobilů v závislosti na výchozí rychlosti. K dispozici je celkem $n = 63$ měření. Proměnná rychlost udává výchozí rychlost (míle/hod.) před začátkem brzdění, proměnná dráha pak udává odpovídající brzdnou dráhu uvedenou ve stopách.

Rychlost	4	5	5	5	5	7	7	8	8	8
Dráha	4	2	8	8	4	6	7	9	8	13
Rychlost	8	9	9	9	10	10	10	12	12	12
Dráha	11	5	13	5	8	17	14	11	21	19
Rychlost	13	13	13	14	14	15	16	16	16	17
Dráha	18	27	15	14	16	16	19	14	34	29
Rychlost	17	18	18	18	19	20	21	21	21	22
Dráha	22	47	29	34	30	48	55	39	42	35
Rychlost	24	25	25	25	25	26	26	27	27	28
Dráha	56	33	59	48	56	39	41	78	57	64
Rychlost	28	29	29	30	30	30	31	35	35	36
Dráha	84	68	54	60	101	67	77	85	107	79
Rychlost	39	40	40							
Dráha	138	110	134							

a) Sestrojte bodový graf.

b) Určete regresní přímku a nakreslete její graf.

c) Určete regresní parabolu a nakreslete její graf.

d) Odhadněte brzdnou dráhu pro rychlost 25 mil/hod.

[Datový soubor: brzdna_draha.txt]

4. V padesátých letech došlo k úniku radioaktivního odpadu ze skládky v Hanfordu ve státě Washington do řeky Columbia River. V devíti okrscích níže po proudu ve státě Oregon bylo počítáno vystavení radioaktivitě X (na základě vzdálenosti od Hanfordu a vzdálenosti průměrného obyvatele od řeky apod.). Současně se sledovala úmrtnost na rakovinu Y (úmrtnost na 100 000 lidí za rok v letech 1959–64). Získané údaje jsou shrnuty v následující tabulce.

okrsek	radioaktivní vystavení X	úmrtnost na rakovinu Y
Clatsop	8,3	210
Columbia	6,4	180
Cilliam	3,4	130
Hood River	3,8	170
Morrow	2,6	130
Portland	11,6	210
Sherman	1,2	120
Umatilla	2,5	150
Wasco	1,6	140

Pro daný datový soubor odhadněte parametry těchto modelů:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
$Y_i = \beta_0 + \beta_1 \ln x_i + \epsilon_i, i = 1, \ldots, n.$

[Datový soubor: radiace_umrtnost.txt]

5. Cílem studie bylo nalézt závislost mezi tělesným tukem lehkých atletů-běžců y, kteří trénují asi 12 hodin, a zkonzumovaným tukem v jejich každodenní stravě x. U náhodného vzorku 18 běžců byl měřen jejich tělesný podkožní tuk y [%] a sledován v závislosti na zkonzumovaném tuku ve stravě x [%].

x	22,0	30,0	24,0	22,0	21,0	36,0	14,0	17,0	20,0
y	9,80	9,70	12,00	11,70	11,60	11,60	8,00	8,60	10,40
x	21,0	35,0	37,0	32,0	35,0	35,0	26,0	24,0	14,0
y	9,70	11,20	10,80	10,90	12,30	11,50	7,80	10,20	7,90

Pro daný datový soubor odhadněte parametry těchto modelů:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\, i = 1, \ldots, n$
$Y_i = \beta_0 + \frac{\beta_1}{x_i} + \epsilon_i,\, i = 1, \ldots, n.$ [Datový soubor: tuk_sportovci.txt]

6. U automobilu byla změřena spotřeba Y v závislosti na rychlosti X. Údaje jsou v tabulce

Rychlost [km/hod.]	40	50	60	70	80	90	100	110
Spotřeba [l/100 km]	5,7	5,4	5,2	5,2	5,8	6,0	7,5	8,1

Pro daný datový soubor odhadněte parametry těchto modelů:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\, i = 1, \ldots, n$
$Y_i = \beta_0 + \beta_1 \ln x_i + \beta_2 x_i^2 + \epsilon_i,\, i = 1, \ldots, n.$ [Datový soubor: rychlost_spotreba.txt]

7. Data popisují výsledky vstupních zdravotních testů uchazečů o službu u policie.

Tlak	66	87	85	59	76	77	70	66	75	66
Hmotnost	87,36	117,6	82,85	62,32	82	102	70,12	88,07	77,96	74,33
Tuk	16,98	27,6	6,61	3,26	19	27	6,88	18,8	18,87	8,15
Tlak	74	68	72	76	94	63	80	67	77	78
Hmotnost	56,2	81,75	80,24	74,81	61,98	95,23	72,48	92,45	104,56	66,2
Tuk	3,44	20,31	12,96	12,42	3,58	12,91	11,34	17,5	18,93	10,94
Tlak	77	67	78	78	80	95	76	78	73	80
Hmotnost	87,16	82,42	64,11	81,57	99,85	78,49	87,13	65,64	51,76	67,14
Tuk	17,72	9,55	9,54	13,1	17,75	9,57	18,52	6,4	2,86	4,31
Tlak	81	61	65	69	66	75	72	66	93	77
Hmotnost	78,74	86,83	70,48	72,67	85,86	84,86	66,97	68,33	63,34	85,72
Tuk	16,26	9,72	6,29	4,37	14,43	17	5,8	8,14	3,63	23,61
Tlak	68	71	84	81	74	79	89	79	80	67
Hmotnost	89	95,17	84,19	63,12	70,01	82,11	71	94,56	70,91	79,19
Tuk	18,83	19,16	15,83	8,77	6,61	22,22	8,29	26,82	9,32	19,9

Popište vhodným regresním modelem (pokud to lze) závislost tlaku na hmotnosti a procenta tuku v těle. Najděte vhodný model pro popis závislosti hmotnosti na procentech tuku v těle.
[Datový soubor: vstupni_testy.txt]