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

Testování hypotéz o parametrech regresního modelu

1 Testy významnosti parametrů

Mějme lineární regresní model (LRM)

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

kde

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

Odhady neznámých parametrů metodou nejmenších čtverců jsou dány

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

reziduální součet čtverců je

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

Předpokládejme nyní, že náhodné chyby $e_i, i = 1\ldots, n$ v lineárním regresním modelu mají normální rozdělení s nulovou střední hodnotou a rozptylem σ^2. Potom mají odhady $\widehat{\beta}_j, j = 1, \ldots, k$ regresní koeficientů β_j normální rozdělení, tedy platí $\widehat{\beta}_j \sim N(\beta_j, D(\widehat{\beta}_j))$, kde rozptyly $D(\widehat{\beta}_j)$ jsou dány:

$$D(\widehat{\beta}_1) = \sigma^2 h_{11}, D(\widehat{\beta}_2) = \sigma^2 h_{22}, \ldots, D(\widehat{\beta}_k) = \sigma^2 h_{kk},$$

přičemž $h_{11}, h_{22}, \ldots, h_{kk}$ jsou prvky na hlavní diagonále matice $\mathbf{H} = (\mathbf{X}'\mathbf{X})^{-1}$. Rozptyly odhadů regresních parametrů odhadneme $\widehat{D}(\widehat{\beta}_j) = s^2 h_{jj}$, druhé odmocniny těchto odhadů

$$s(\widehat{\beta}_j) = \sqrt{s^2 h_{jj}}$$

se nazývají **směrodatné chyby** odhadů regresních parametrů. Testy významnosti parametrů β_j, $j = 1, \ldots, k$ (jejich nenulovosti) jsou založeny na statistikách

$$T = \frac{\widehat{\beta}_j - \beta_j}{s(\widehat{\beta}_j)},$$

které mají Studentovo rozdělení s $n - k$ stupni volnosti.

Budeme testovat nulovou hypotézu H: $\beta_j = 0$ proti alternativní hypotéze A: $\beta_j \neq 0$. Při platnosti nulové hypotézy má statistika

$$T = \frac{\widehat{\beta}_j - \beta_j}{s(\widehat{\beta}_j)} = \frac{\widehat{\beta}_j - 0}{s(\widehat{\beta}_j)} = \frac{\widehat{\beta}_j}{s(\widehat{\beta}_j)}$$

Studentovo rozdělení s $n - k$ stupni volnosti. Kritickou hodnotou odpovídající hladině významnosti α je tedy kvantil $t_{1-\frac{\alpha}{2}}(n - k)$.

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.

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

2 Test významnosti modelu

Zřejmě platí, že $y_i - \overline{y} = (y_i - \widehat{y}_i) + (\widehat{y}_i - \overline{y})$. Lze ukázat, že také platí

$$\sum_{i=1}^{n}(y_i-\overline{y})^2 = \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 + \sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2 \to S_Y = S_e + S_T,$$

kde

- **celkový součet čtverců** $S_Y = \mathbf{Y}'\mathbf{Y} - n\overline{y}^2$

$$S_Y = \sum_{i=1}^{n}(y_i-\overline{y})^2 = n\cdot s^2(y),\ \text{kde}\ s^2(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2$$

- **reziduální součet čtverců** $S_e = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$

$$S_e = \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 = (n-k)\cdot s^2(y),\ \text{kde}\ s_e^2(y) = \frac{1}{n-k}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2$$

- **teoretický součet čtverců** $S_T = \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\overline{y}^2$

$$S_T = \sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2 = n\cdot s^2(\hat{y}),\ \text{kde}\ s^2(\widehat{y}) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i-\overline{y})^2$$

Pro regresní přímku $\widehat{y} = \widehat{\beta}_1 + \widehat{\beta}_2 x$ dostáváme

$$\begin{aligned}
S_e &= \sum_{i=1}^{n}(y_i-\widehat{y}_i)^2 = \sum_{i=1}^{n}(y_i-\widehat{\beta}_1-\widehat{\beta}_2x_i)^2 = \dots = \\
&= \sum_{i=1}^{n}y_i^2 - \widehat{\beta}_1\sum_{i=1}^{n}y_i - \widehat{\beta}_2\sum_{i=1}^{n}x_iy_i \\
S_T &= \sum_{i=1}^{n}(\widehat{y}_i-\hat{y}_i)^2 = \sum_{i=1}^{n}\left(\widehat{\beta}_1+\widehat{\beta}_2x_i-\frac{1}{n}\sum_{i=1}^{n}y_i^2\right) = \dots = \\
&= \widehat{\beta}_1\sum_{i=1}^{n}y_i + \widehat{\beta}_2\sum_{i=1}^{n}x_iy_i - \frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2 \\
S_Y &= S_R + S_T = \dots = \sum_{i=1}^{n}y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2
\end{aligned}$$

- teoretický součet čtverců S_T je ta část celkového součtu čtverců S_Y, která je vysvětlená zvolenou regresní funkcí
- reziduální součet čtverců S_e je ta část celkového součtu čtverců S_Y, která zvolenou regresní funkcí vysvětlená není

Při ověřování významnosti regresního modelu (obvykle označovaný jako F-test modelu) se testuje nulová hypotéza H: $\beta_1 = c, c \neq 0, \beta_2 = \beta_3 = \cdots = \beta_k = 0$ proti alternativní hypotéze A: $\beta_j \neq 0$ pro alespoň jedno $j = 2, 3, \ldots, k$. Testové kritérium je statistika

$$F = \frac{S_T(y)}{k-1} : \frac{S_e(y)}{n-k},$$

která má při platnosti nulové hypotézy Fisher-Snedecorovo rozdělení F s $k-1$ a $n-k$ stupni volnosti, Kritickou hodnotou je kvantil $F_{1-\alpha}(k-1, n-k)$ daného F rozdělení.

- Jsou-li celkový F-test i všechny t-testy jsou statisticky významné, model se považuje za vhodný k vystižení variability proměnné Y (to však ještě neznamená, že je model správně navržen).
- Jsou-li celkový F-test i všechny t-testy jsou statisticky nevýznamné, model se považuje za nevhodný, protože nevystihuje variabilitu proměnné Y.
- Je-li celkový F-test statisticky významný, ale některé t-testy vychází nevýznamné, model se považuje za vhodný, ale provádí se zpravidla vypuštění nevýznamných parametrů.
- Je-li celkový F-test statisticky významný, ale všechny t-testy vychází nevýznamné – paradox: formálně model jako celek vyhovuje, ale žádný člen modelu sám o sobě významný není – jde o důsledek tzv. **multikolinearity**, tj. lineární závislosti mezi jednotlivými regresory.

Vhodnost zvoleného modelu lze vyjádřit pomocí tzv. indexu (koeficientu) determinace, který je definován jako podíl variability, kterou je schopen popsat regresní model, ku celkové variabilitě vysvětlované proměnné $S_c(y) = \sum_{i=1}^{n}(y_i - \overline{y_i})^2$

$$R^2 = \frac{S_T(y)}{S_c(y)} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}.$$

Toto číslo nabývá hodnot z intervalu $\langle 0, 1 \rangle$. Čím více se R^2 blíží k 1, tím považujeme danou závislost za silnější, a tedy dobře vystiženou použitým regresním modelem; naopak čím více se bude blížit k 0, tím považujeme danou závislost za slabší a regresní funkci za méně výstižnou. Nízká hodnota R^2 ještě nemusí znamenat nízký stupeň závislosti mezi proměnnými, ale může signalizovat chybnou volbu regresního modelu. R^2 představuje výběrový index determinace, který lze použít jako odhad teoretického indexu determinace. Tento odhad je asymptoticky nestranný, nicméně pro malé výběry nadhodnocuje skutečnou těsnost závislosti a je závislý na počtu parametrů regresního modelu. Lze provést jeho korekci

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-k},$$

čím získáme odhad nestranný.

3 Test obecné lineární hypotézy

Z předchozích výsledků lze za předpokladu, že LRM je plné hodnosti, snadno odvodit (viz Zvára (2008)), že pro libovolnou reálnou matici $\mathbf{A}$ typu $m \times k$ a hodnosti $m \leq k$ má statistika

$$F = \frac{1}{ms^2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{A}'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}\mathbf{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Fisher-Snedecorovo $F-$rozdělení o m a $n-k$ stupních volnosti. Tuto statistiku lze pak využít při testování obecné lineární hypotézy H_0, kterou zapíšeme ve tvaru

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{a}, \tag{1}$$

kde $\mathbf{a}$ je vhodný $m-$rozměrný reálný vektor, pro něž je rovnice (1) řešitelná. Odtud pak plyne, že testovací statistika

$$F = \frac{1}{ms^2}(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{a})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{a}) \tag{2}$$

má za platnosti nulové hypotézy H_0 Fisher-Snedecorovo F rozdělení o m a $n-k$ stupních volnosti. Proto nulovou hypotézu H_0 zamítáme na hladině významnosti α, když $F > F_{1-\alpha}(m, n-k)$, kde $F_{1-\alpha}(m, n-k)$ je $(1-\alpha)-$kvantil $F-$rozdělení o m a $n-k$ stupních volnosti. Speciální volbou matice $\mathbf{A}$ a vektoru $\mathbf{a}$ v (1) lze potom získat speciální hypotézy, které experimentátora zajímají. Po dosazení za $\mathbf{A}$ a $\mathbf{a}$ do statistiky F ve (2) lze získat odpovídající testovací statistiky pro testování hypotéz o parametrech $\beta_1, \ldots, \beta_k$.

Tímto způsobem lze konstruovat řadu běžných testů o neznámých parametrech včetně testů o parametrech LRM, které byly uvedeny dříve.

Příklad 1. Test hypotézy $\beta_j = 0$ lze získat při volbě $\mathbf{A} = \mathbf{u}_j$, kde $\mathbf{u}_j$ je $k-$rozměrný jednotkový vektor s jedničkou na $j-$tém místě a $\mathbf{a} = 0$. Jde o test, založený na testovací statistice T.

Příklad 2. Testy hypotéz o rovnosti parametrů, např. test hypotézy $\beta_1 = \beta_2 = \cdots = \beta_m$, $m \leq k$, lze získat volbou $\mathbf{a} = \mathbf{0}_{m-1}$ a

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}$$

je matice typu $(m-1) \times k$.

Příklad 3. Předpokládejme, že jsou dány dva nezávislé regresní modely, každý s jedním regresorem. První o rovnici $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ a druhý o rovnici $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{e}^*$, kde $X_{n\times 2}$ a $X^*_{n^*\times 2}$ jsou příslušné matice plánů s vektorem jedniček v prvním sloupci, $\mathbf{Y}$ a $\mathbf{Y}^*$ jsou vektory pozorování závisle proměnných v obou modelech, $\mathbf{e}$ a $\mathbf{e}^*$ jsou nezávislé normálně rozdělené vektory náhodných chyb a konečně $\boldsymbol{\beta}$ a $\boldsymbol{\beta}^*$ jsou dvourozměrné vektory neznámých parametrů. Tedy jsou dány dva nezávislé regresní modely a regresní funkce v obou modelech je dána přímkou. Když oba vektory náhodných chyb mají varianční matice $\sigma^2\mathbf{I}_n$ a $\sigma^2\mathbf{I}_{n^*}$, lze vytvořit spojení obou modelů a uvážit nový model tvaru

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^* \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} \mathbf{e} \\ \mathbf{e}^* \end{pmatrix}$$

V tomto modelu lze např. testovací statistiku pro testování rovnoběžnosti obou regresních přímek snadno dostat ze vzorce (2) volbou $\mathbf{A} = (0, 1, 0, -1)$, $\mathbf{a} = 0$.

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) Určete odhady parametrů regresní přímky, testujte významnost regresních parametrů na hladině významnosti 0,05.

   b) Proveďte test významnosti regresního modelu, určete koeficient determinace.

[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) Určete odhady parametrů regresní přímky, testujte významnost těchto parametrů na hladině významnosti 0,05..

   b) Určete bodové odhady regresní logaritmické křivky, testujte významnost těchto parametrů.

   c) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[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) Určete odhady parametrů regresní přímky, testujte významnost těchto parametrů na hladině významnosti 0,05..

b) Určete odhady parametrů kvadratické regresní funkce, testujte významnost těchto parametrů.

c) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[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:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
$Y_i = \beta_0 + \beta_1 \ln x_i + \epsilon_i, i = 1, \ldots, n.$

a) Určete odhady parametrů regresní přímky, testujte významnost těchto parametrů na hladině významnosti 0,05.

b) Určete odhady parametrů logaritmické regresní funkce, testujte významnost těchto parametrů.

c) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[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:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
$Y_i = \beta_0 + \frac{\beta_1}{x_i} + \epsilon_i, i = 1, \ldots, n.$

a) Pro regresní přímku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
b) Pro hyperbolickou regresní křivku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
c) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[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:
$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n$
$Y_i = \beta_0 + \beta_1 \ln x_i + \beta_2 x_i^2 + \epsilon_i, i = 1, \ldots, n.$

a) Pro regresní přímku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
b) Pro logaritmickou regresní křivku testujte významnost regresních parametrů na hladině významnosti 0,05 a 0,01.
c) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodnější, zdůvodněte.

[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

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

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.

a) Testujte významnost regresních parametrů pro oba modely na hladině významnosti 0,05.
b) Pro oba modely poveďte test významnosti regresního modelu, určete koeficienty determinace. Rozhodněte, který z uvedených modelů je pro popis studované závislosti vhodný, případně který nikoli. Své závěry zdůvodněte.

[Datový soubor: vstupni_testy.txt]