

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

**VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA
FAKULTA STROJNÍ**



MATEMATIKA III – V PŘÍKLADECH

Cvičení 10 – Regrese a korelace

Mgr. Petr Otipka

Ostrava 2013

© Mgr. Petr Otipka

© Vysoká škola báňská – Technická univerzita Ostrava

ISBN 978-80-248-3034-6



Tento studijní materiál vznikl za finanční podpory Evropského sociálního fondu (ESF) a rozpočtu České republiky v rámci řešení projektu: CZ.1.07/2.2.00/15.0463, MODERNIZACE VÝUKOVÝCH MATERIÁLŮ A DIDAKTICKÝCH METOD

OBSAH

10	REGRESE A KORELACE	3
10.1	Řešené úlohy	3
10.1.1	Úlohy k řešení	3
10.1.2	Výsledky úloh k řešení	4
10.1.3	Sada testovacích otázek.....	5
10.1.4	Správné odpovědi k testovacím otázkám	5



10 REGRESE A KORELACE

10.1 ŘEŠENÉ ÚLOHY

10.1.1 Úlohy k řešení

1.1. Charakterizujte závislost proměnné y na x regresní funkcí ve tvaru:

a) $Y = a + \frac{b}{x}$

b) $Y = ax^2 + bx + c$

Určete indexy korelace

x	1	1	3	4	6
y	0	1	4	5	5

1.2. Při seskoku parašutisty byla měřena závislost mezi rychlostí v [m/s] a tlakem p [0,1mPa] na povrchu padáku. Výsledky vyrovnejte parabolou $p = a + bv^2$. Vypočtěte index korelace.

v	2,4	3,5	5	6,89	10
p	0,0141	0,0281	0,0562	0,1125	0,225

1.3. Charakterizujte těsnost zvolené závislosti ve tvaru $Y = a + b \cdot \log x$ mezi proměnnými x a y . Vypočtěte index korelace.

x	1	1	3	3	5	6	7	7
y	70	104	162	210	200	250	240	260

1.4. Zjišťovalo se, zda u souboru chlapců je závislost v počtu provedených shybů a kliků. Výsledky jsou zaznamenány v tabulce:

chlapec	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
počet shybů	1	3	2	0	5	6	1	4	3	5	6	2	1	1	8
počet kliků	10	15	15	0	40	25	7	31	30	35	41	10	14	9	64

a) Určete, zda je mezi počtem shybů a počtem kliků silná lineární závislost, určete její míru.

b) Najděte nejvhodnější regresní funkci závislosti mezi počtem shybů a kliků.

1.5. V tabulce jsou průměrné doby trvání slunečního svitu za období 1961–1990 v Praze-Ruzyni (argument x) a dlouhodobé průměrné teploty vzduchu za období 1961–1990 v Praze-Ruzyni (argument y).

	leden	únor	březen	duben	květen	červen	červenec	srpen	září	říjen	listopad	prosinec
x [h]	50,0	72,4	124,7	167,6	214,0	218,6	226,7	212,3	161,0	120,8	53,6	46,7
y [°C]	-2,4	-0,9	3,0	7,7	12,7	15,9	17,5	17,0	13,3	8,3	2,9	-0,6

Najděte vhodnou regresní funkci závislosti y na x a vypočtěte index determinace.

1.6. Tabulka obsahuje data o vývoji indexu spotřebitelských cen – životních nákladů v ČR.

průměr roku 2005 = 100

rok	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
index	59,1	64,5	70,2	76,2	84,4	86,2	89,4	93,6	95,4	95,5	98,1	100,0	102,5	105,4	112,1	113,3	114,9

Proložte daty vhodnou regresní funkcí a využijte ji k odhadu hodnoty indexu v roce 2015.



- 1.7. Podobně jako v předchozím případě – z údajů v tabulce průměrných mezd v ČR odhadněte průměrný plat v roce 2015.

rok	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
mzda	13 219	14 378	15 524	16 430	17 466	18 344	19 546	20 957	22 592	23 344	23 797

- 1.8. Popište závislost produktivity práce (v desítkách kusů za hodinu) na % plnění norem x a průměrném věku pracovníků z v pěti firmách, zabývajících se výrobou stejného druhu výrobků pomocí regresní roviny. Vypočítejte hodnotu indexu korelace a odhadněte hodnotu produktivity práce při plnění norem na 106 % a při průměrném věku pracovníků 32 let.

y - produktivita	1	2	2	3	3
x - plnění normy	0,9	1,02	1,05	1,3	1,4
z - průměrný věk	30	27	26	27	25

- 1.9. V tabulce jsou uvedeny hodnoty tří nezávisle proměnných x_1 , x_2 a x_3 a hodnota závisle proměnné y . Stanovte lineární regresi mezi těmito veličinami a hodnotu indexu determinace.

y	11	7	10	11	10	9	15	16
x_1	1	-1	0	4	2	4	1	2
x_2	1	2	3	-2	1	3	-1	0
x_3	1	0	2	-3	0	-1	1	2

- 1.10. Posuďte vliv jednotlivých vybraných ukazatelů parních elektráren v roce 1984 na měrné náklady elektráren. Úlohu řešte vícenásobnou lineární regresní analýzou.

elektrárna	y	x_1	x_2	x_3	x_4
Mělník 2	249	0,95	6,86	14,01	12,92
Počerady 1	203	2,27	7,56	12,06	11,74
Chvaletice	256	2,34	6,79	15,03	11,74
Dětmárovice	306	4,34	7,25	17,38	11,7
Tušimice 1	227	2,22	6,58	10,28	12,49
Tušimice 2	213	2,62	7,35	10,12	12,13
Pruněřov 1	349	5,18	6,66	11,26	13,49
Pruněřov 2	210	4,24	7,47	11,53	11,15

y ... měrné náklady [Kč/MWh], x_1 ... poruchy [%], x_2 ... využití pohotového výkonu [tisíce hodin], x_3 ... cena paliva [Kč/GJ], x_4 ... měrná spotřeba [GJ/MWh]

10.1.2 Výsledky úloh k řešení

1.1. a) $Y = 6,06 - \frac{5,565}{x}$; $I = 0,985$; b) $Y = -2,15 + 2,942x - 0,2913x^2$; $I = 0,99$

1.2. $p = 0,00144 + 0,0022506v^2$; $I = 0,9996$

1.3. $Y = 88,32 + 191,54 \log x$; $I = 0,96$

1.4. Lineární funkce: $Y = 6,6939x + 1,6463$; $I = 0,927577$

Kvadratická funkce: $Y = 0,243x^2 + 4,8667x + 3,7354$; $I = 0,93043$

1.5. $Y = 0,098x - 5,766$; $I^2 = 0,888$

1.6. $\text{index}_{2015} = 133,8$

1.7. $\text{mzda}_{2015} = 29699$

1.8. $Y = 3,2456x - 0,0905z + 0,9640$; 1,5071

1.9. $Y = 0,9799x_1 - 1,2680x_2 + 1,8424x_3 + 10,1816$; $I^2 = 0,9636$



$$1.10. \quad Y = -317,561 + 22,295x_1 - 19,545x_2 + 1,187x_3 + 41,457x_4, \quad I^2 = 0,9908$$

10.1.3 Sada testovacích otázek

T10.1. Index korelace používáme k

- a) posouzení míry lineární závislosti mezi veličinami.
- b) vyjádření závislosti mezi dvěma statistickými znaky.
- c) posouzení vhodnosti regresní funkce.

T10.2. Regresí rozumíme

- a) míru (intenzitu) závislosti statistických znaků.
- b) vystižení charakteru závislosti mezi veličinami a určení její konkrétní formy.
- c) vypočtení konstant a, b z rovnice $Y = ax + b$.

T10.3. Metodu nejmenších čtverců odchylek používáme ve statistice k

- a) nalezení nejlepšího funkčního předpisu statistické závislosti.
- b) vypočtení koeficientu korelace.
- c) vypočtení indexu korelace.

T10.4. Koeficient korelace používáme k

- a) posouzení míry lineární závislosti mezi veličinami.
- b) vyjádření závislosti mezi dvěma statistickými znaky.
- c) posouzení vhodnosti regresní funkce.

T10.5. Index korelace může nabývat v případě lineární regrese hodnot z intervalu

- a) $\langle -1, 1 \rangle$
- b) $\langle 0, 1 \rangle$
- c) $\langle 0, \infty \rangle$

T10.6. Při různých rychlostech byly naměřeny následující hodnoty spotřeby automobilu:

rychlost x	50	60	70	80	90	100	110	120	130
spotřeba y	8,5	7,8	7,3	7,1	6,9	7	7,4	7,9	8,6

Která regresní křivka bude nejlépe vystihovat závislost mezi rychlostí x a spotřebou y (hodnoty a, b , případně c by se vypočetly například metodou nejmenších čtverců)?

- a) $Y = a + bx$
- b) $Y = a + bx + cx^2$
- c) $Y = a + b \cdot \log x$

T10.7. Blíží-li se hodnota indexu korelace k nule, pak

- a) se mezi danými veličinami jedná vždy o slabou lineární závislost.
- b) se mezi danými veličinami jedná vždy o silnou lineární závislost.
- c) vypočtená regresní funkce nevystihuje statistickou závislost mezi veličinami.

10.1.4 Správné odpovědi k testovacím otázkám

T10.1. c)

T10.2. b)

T10.3. a)

T10.4. a)

T10.5. b)

T10.6. b)

T10.7. c)

