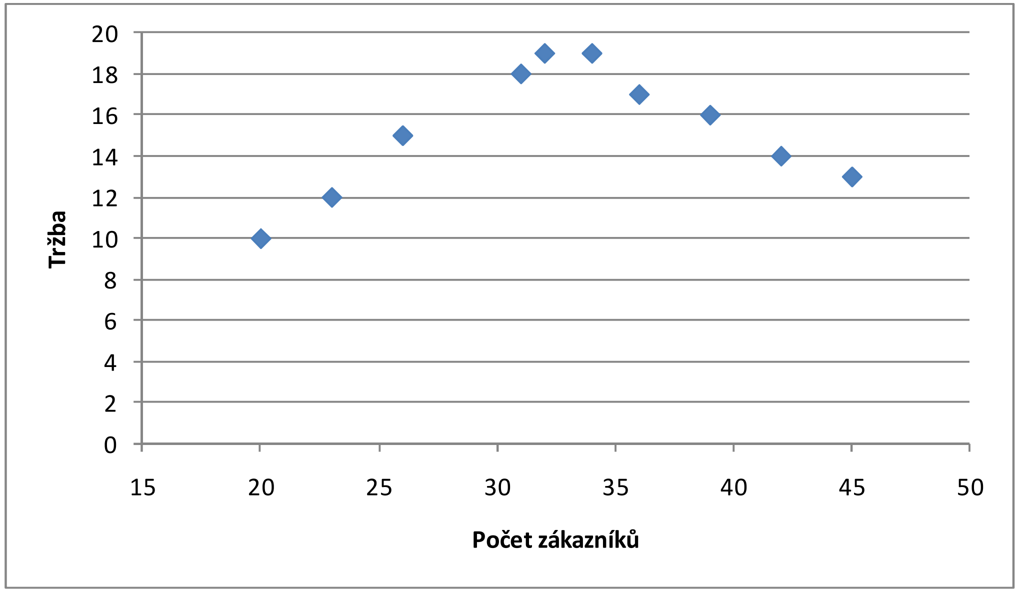
Regresní analýza

Regresní analýza

- **Př.1:** V obchodě, zabývajícím se prodejem náhradních dílů do automobilů, bylo provedeno měření počtu zákazníků přicházejících do obchodu za 1 hodinu a odpovídajících tržeb za 1 hodinu vyjádřených v tisících Kč. Stanovte Pearsonův korelační koeficient a vyberte vhodný regresní model vyjadřující závislost hodinové tržby na počtu přicházejících zákazníků.

Regresní analýza

Počet zákazníků - x_i	Hodinová tržba - Y_i
20	10
23	12
26	15
31	18
32	19
34	19
36	17
39	16
42	14
45	13



Regresní analýza

$$r_{x,Y} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$

	x_i	Y_i	$(x_i\text{-}x_p)$	$(Y_i\text{-}Y_p)$	$(x_i\text{-}x_p)\cdot(Y_i\text{-}Y_p)$	$(x_i\text{-}x_p)^2$	$(Y_i\text{-}Y_p)^2$
	20	10	-12,80	-5,30	67,84	163,84	28,09
	23	12	-9,80	-3,30	32,34	96,04	10,89
	26	15	-6,80	-0,30	2,04	46,24	0,09
	31	18	-1,80	2,70	-4,86	3,24	7,29
	32	19	-0,80	3,70	-2,96	0,64	13,69
	34	19	1,20	3,70	4,44	1,44	13,69
	36	17	3,20	1,70	5,44	10,24	2,89
	39	16	6,20	0,70	4,34	38,44	0,49
	42	14	9,20	-1,30	-11,96	84,64	1,69
	45	13	12,20	-2,30	-28,06	148,84	5,29
Součet	328	153			68,60	593,60	84,10
Průměr	32,80	15,30					

$r_{x,Y}$
0,30703
CORREL
0,30703

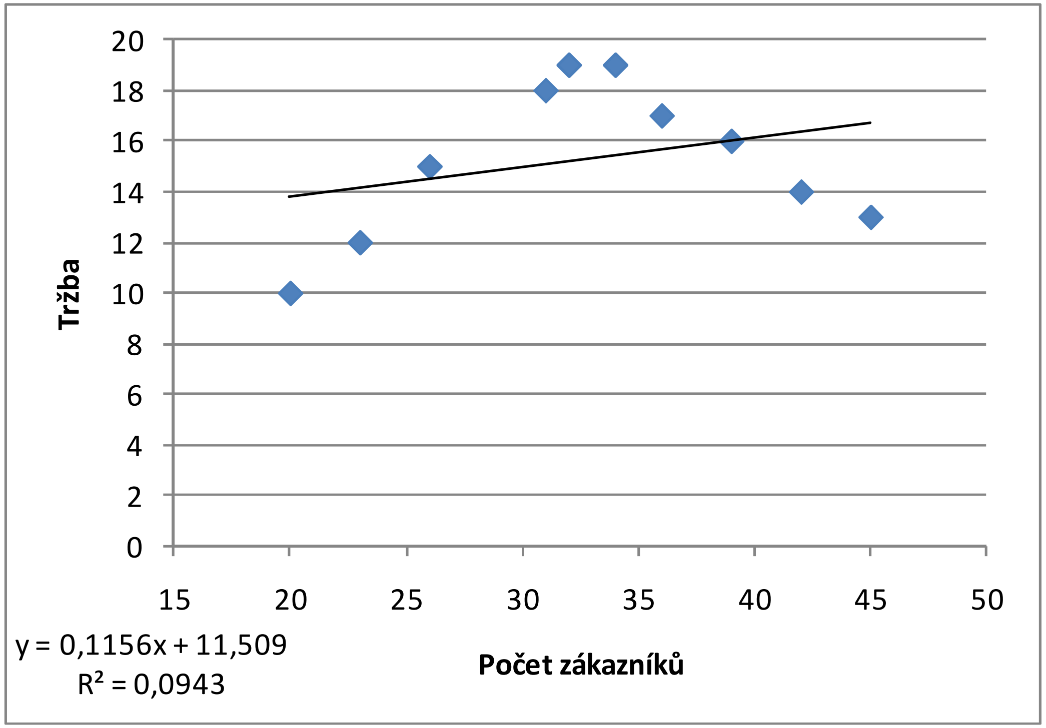
Regresní analýza

1) Vyrovnávací křivka bude přímka:

$$\hat{Y}_i = b_0 + b_1 \cdot x_1 ,$$

$$\text{kde } b_0 = \overline{Y} - b_1 \overline{x} ,$$

$$b_1 = \frac{\sum_{i=1}^{n} Y_i \cdot (x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} .$$



Regresní analýza

	x_i	Y_i	x_i-x_p	$Y_i\cdot(x_i$-$x_p)$	$(x_i$-$x_p)^2$
	20	10	-12,80	-128,00	163,84
	23	12	-9,80	-117,60	96,04
	26	15	-6,80	-102,00	46,24
	31	18	-1,80	-32,40	3,24
	32	19	-0,80	-15,20	0,64
	34	19	1,20	22,80	1,44
	36	17	3,20	54,40	10,24
	39	16	6,20	99,20	38,44
	42	14	9,20	128,80	84,64
	45	13	12,20	158,60	148,84
Součet	328	153		68,60	593,60
Průměr	32,80	15,30			

b_0	b_1
11,51	0,12

$$\hat{Y}_i = 11{,}51 + 0{,}12 x_1$$

Pozn.

$$x_p = \overline{x}$$

$$Y_p = \overline{Y}$$

Regresní analýza

	x_i	Y_i	$\hat{Y}_i$	$(\hat{Y}_i - Y_p)^2$	$(Y_i - Y_p)^2$
	20	10	13,82	2,19	28,09
	23	12	14,17	1,28	10,89
	26	15	14,51	0,62	0,09
	31	18	15,09	0,04	7,29
	32	19	15,21	0,01	13,69
	34	19	15,44	0,02	13,69
	36	17	15,67	0,14	2,89
	39	16	16,02	0,51	0,49
	42	14	16,36	1,13	1,69
	45	13	16,71	1,99	5,29
Součet	328	153		7,93	84,10
Průměr	32,80	15,30			

R^2
0,09

Regresní analýza

2) Vyrovnávací křivka bude parabola:

$$\widehat{Y}_i = b_0 + b_1 \cdot x_i + b_2 \cdot x_i^2 ,$$

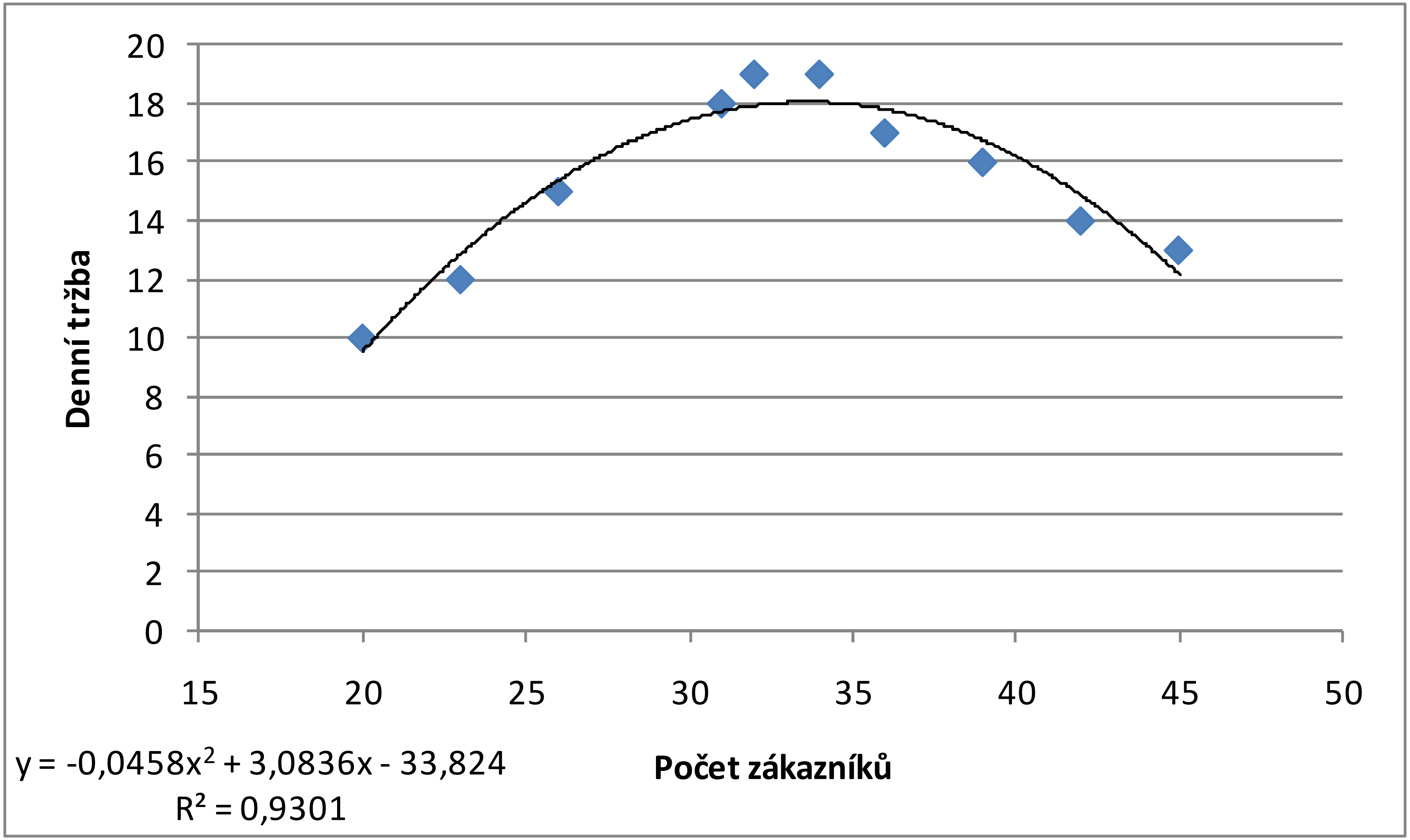
kde odhady koeficientů modelu získáme řešením soustavy:

$$\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} x_i + b_2 \sum_{i=1}^{n} x_i^2 ,$$

$$\sum_{i=1}^{n} Y_i \cdot x_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 + b_2 \sum_{i=1}^{n} x_i^3 ,$$

$$\sum_{i=1}^{n} Y_i \cdot x_i^2 = b_0 \sum_{i=1}^{n} x_i^2 + b_1 \sum_{i=1}^{n} x_i^3 + b_2 \sum_{i=1}^{n} x_i^4 .$$

Regresní analýza



Regresní analýza

	x_i	Y_i	x_i^2	x_i^3	x_i^4	$Y_i \cdot x_i$	$Y_i \cdot x_i^2$
	20	10	400	8000	160000	200	4000
	23	12	529	12167	279841	276	6348
	26	15	676	17576	456976	390	10140
	31	18	961	29791	923521	558	17298
	32	19	1024	32768	1048576	608	19456
	34	19	1156	39304	1336336	646	21964
	36	17	1296	46656	1679616	612	22032
	39	16	1521	59319	2313441	624	24336
	42	14	1764	74088	3111696	588	24696
	45	13	2025	91125	4100625	585	26325
Součet	328	153	11352	410794	15410628	5087	176595
Průměr	32,80	15,30					

Regresní analýza

- Nyní je nutno řešit soustavu:

$$153 = 10b_0 + 328b_1 + 11352b_2,$$

$$5087 = 328b_0 + 11352b_1 + 410794b_2,$$

$$176595 = 11352b_0 + 410794b_1 + 15410628b_2.$$

10	328	11352	153
328	11352	410794	5087
11352	410794	15410628	176595

10	328	11352	153
0	-5936	-384484	-686
0	-384484	-25238376	-29094

10	328	11352	153
0	-5936	-384484	-686
0	0	1987053680	-91054040

b_0	b_1	b_2
-33,82	**3,08**	**-0,05**

$$\widehat{Y}_i = -33{,}82 + 3{,}08x_i - 0{,}05x_i^2$$

Regresní analýza

	x_i	Y_i	$\hat{Y}_i$	$(\hat{Y}_i-Y_p)^2$	$(Y_i-Y_p)^2$
	20	10	9,52	33,42	28,09
	23	12	12,86	5,96	10,89
	26	15	15,37	0,01	0,09
	31	18	17,73	5,91	7,29
	32	19	17,93	6,91	13,69
	34	19	18,05	7,55	13,69
	36	17	17,80	6,25	2,89
	39	16	16,74	2,07	0,49
	42	14	14,86	0,20	1,69
	45	13	12,15	9,94	5,29
Součet	328	153		78,22	84,10
Průměr	32,80	15,30			

R^2
0,93

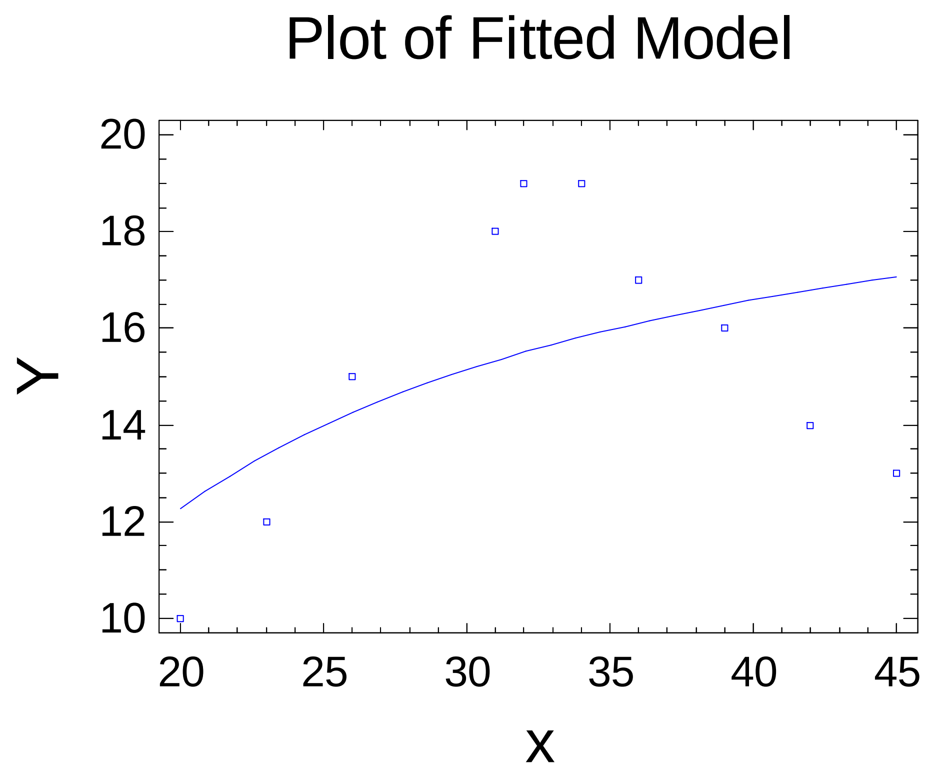
Regresní analýza

3) Vyrovnávací křivka bude hyperbola:

$$\widehat{Y_i} = b_0 + \frac{b_1}{x_i},$$

kde $b_0 = \dfrac{\sum_{i=1}^{n} Y_i \cdot \sum_{i=1}^{n} \frac{1}{x_i^2} - \sum_{i=1}^{n} \frac{Y_i}{x_i} \cdot \sum_{i=1}^{n} \frac{1}{x_i}}{n\sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2}$ a $b_1 = \dfrac{n\sum_{i=1}^{n} \frac{Y_i}{x_i} - \sum_{i=1}^{n} Y_i \cdot \sum_{i=1}^{n} \frac{1}{x_i}}{n\sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^2}$.

Regresní analýza



Y = 20,9162 - 173,173/x   R-squared = 25,814 percent

Regresní analýza

	x_i	Y_i	$1/x_i$	$1/x_i^2$	Y_i/x_i
	20	10	0,05000	0,00250	0,50000
	23	12	0,04348	0,00189	0,52174
	26	15	0,03846	0,00148	0,57692
	31	18	0,03226	0,00104	0,58065
	32	19	0,03125	0,00098	0,59375
	34	19	0,02941	0,00087	0,55882
	36	17	0,02778	0,00077	0,47222
	39	16	0,02564	0,00066	0,41026
	42	14	0,02381	0,00057	0,33333
	45	13	0,02222	0,00049	0,28889
Součet	328	153	0,32431	0,01124	4,83658

b_0	b_1
20,92	**-173,17**

$$\widehat{Y}_i = 20{,}92 - \frac{173{,}17}{x_i}$$

Regresní analýza

	x_i	Y_i	$\hat{Y}_i$	$(\hat{Y}_i-Y_p)^2$	$(Y_i-Y_p)^2$
	20	10	12,26	9,26	28,09
	23	12	13,39	3,66	10,89
	26	15	14,26	1,09	0,09
	31	18	15,33	0,00	7,29
	32	19	15,50	0,04	13,69
	34	19	15,82	0,27	13,69
	36	17	16,11	0,65	2,89
	39	16	16,48	1,38	0,49
	42	14	16,79	2,23	1,69
	45	13	17,07	3,13	5,29
Součet	328	153		21,71	84,10
Průměr	32,80	15,30			

R^2
0,26

Regresní analýza

4) Vyrovnávací křivka bude logaritmická křivka:

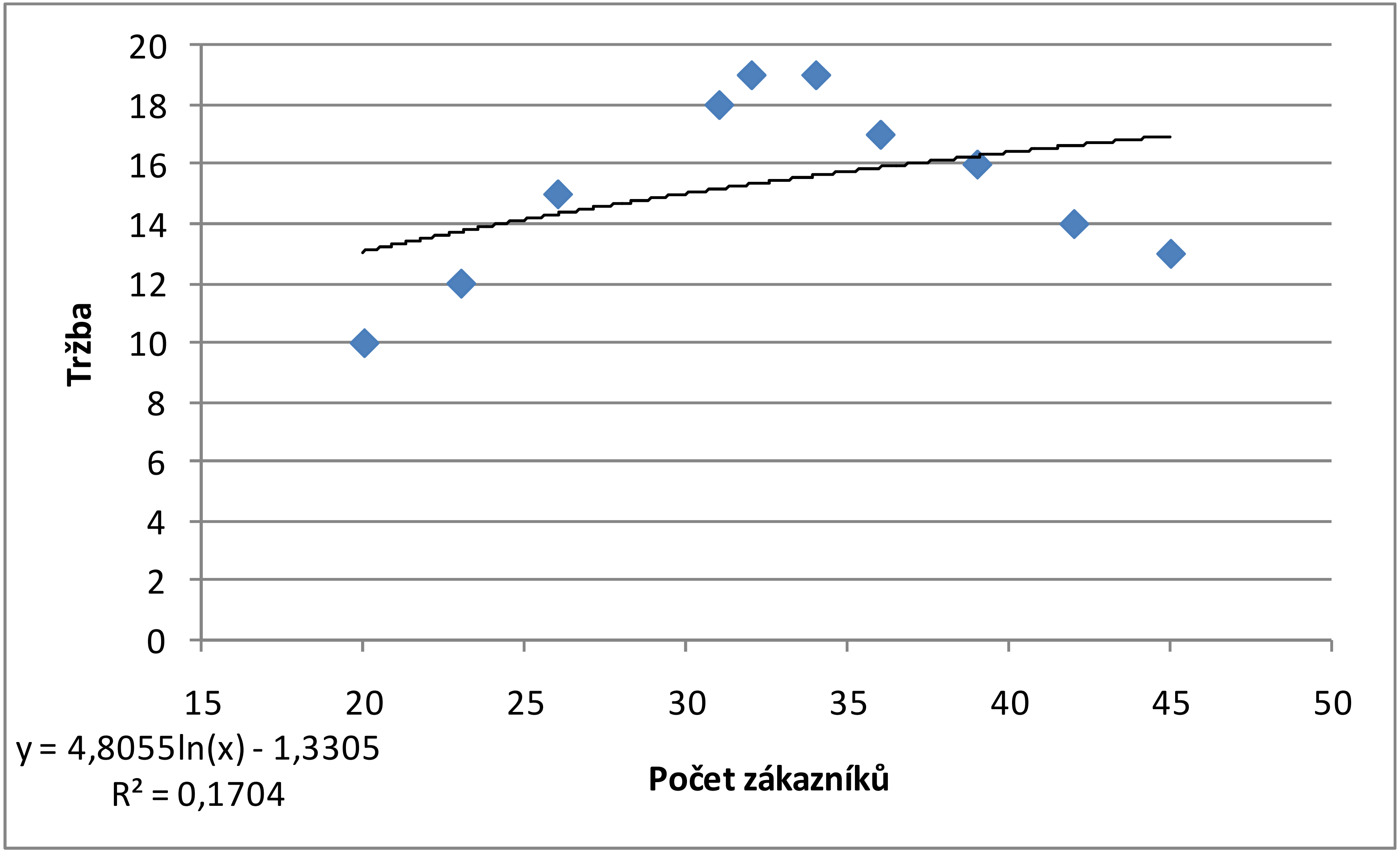
$$\widehat{Y}_i = b_0 + b_1 \cdot \ln x_i \,,$$

$$\text{kde } b_0 = \frac{\sum_{i=1}^{n} Y_i \cdot \ln^2 x_i - \sum_{i=1}^{n} Y_i \cdot \ln x_i \cdot \sum_{i=1}^{n} \ln x_i}{n \sum_{i=1}^{n} \ln^2 x_i - \left(\sum_{i=1}^{n} \ln x_i \right)^2},$$

$$b_1 = \frac{n \sum_{i=1}^{n} Y_i \cdot \ln x_i - \sum_{i=1}^{n} \ln x_i \cdot \sum_{i=1}^{n} Y_i}{n \sum_{i=1}^{n} \ln^2 x_i - \left(\sum_{i=1}^{n} \ln x_i \right)^2}.$$

Pozn. Jelikož software Excel v tomto případě pracuje s přirozeným logaritmem, budeme ho v tomto případě taky používat.

Regresní analýza



Regresní analýza

	x_i	Y_i	$\ln x_i$	$\ln^2 x_i$	$Y_i \cdot \ln x_i$
	20	10	2,99573	8,97441	29,95732
	23	12	3,13549	9,83132	37,62593
	26	15	3,25810	10,61519	48,87145
	31	18	3,43399	11,79227	61,81177
	32	19	3,46574	12,01133	65,84898
	34	19	3,52636	12,43522	67,00085
	36	17	3,58352	12,84161	60,91982
	39	16	3,66356	13,42168	58,61699
	42	14	3,73767	13,97017	52,32737
	45	13	3,80666	14,49068	49,48661
Součet	328	153	34,60682	120,38389	532,46710

b_0	b_1
-1,33	4,81

$$\widehat{Y}_i = -1{,}33 + 4{,}81 \ln x_i$$

Regresní analýza

	x_i	Y_i	$\hat{Y}_i$	$(\hat{Y}_i - Y_p)^2$	$(Y_i - Y_p)^2$
	20	10	13,07	4,99	28,09
	23	12	13,74	2,44	10,89
	26	15	14,33	0,95	0,09
	31	18	15,17	0,02	7,29
	32	19	15,32	0,00	13,69
	34	19	15,62	0,10	13,69
	36	17	15,89	0,35	2,89
	39	16	16,27	0,95	0,49
	42	14	16,63	1,77	1,69
	45	13	16,96	2,76	5,29
Součet	328	153		14,33	84,10
Průměr	32,80	15,30			

R^2
0,17

Regresní analýza

5) Vyrovnávací křivka bude exponenciála:

$$\widehat{Y}_i = b_0 \cdot b_1^{x_i}.$$

Excel pracuje s exponenciální regresní funkcí ve tvaru:

$$\widehat{Y}_i = b_0 \cdot e^{b_1' x_i}.$$

Jelikož platí:

$$b_1 = e^{b_1'} \text{ a } a^y = x \Rightarrow \log_a x = y,$$

potom $b_1' = \ln b_1$.

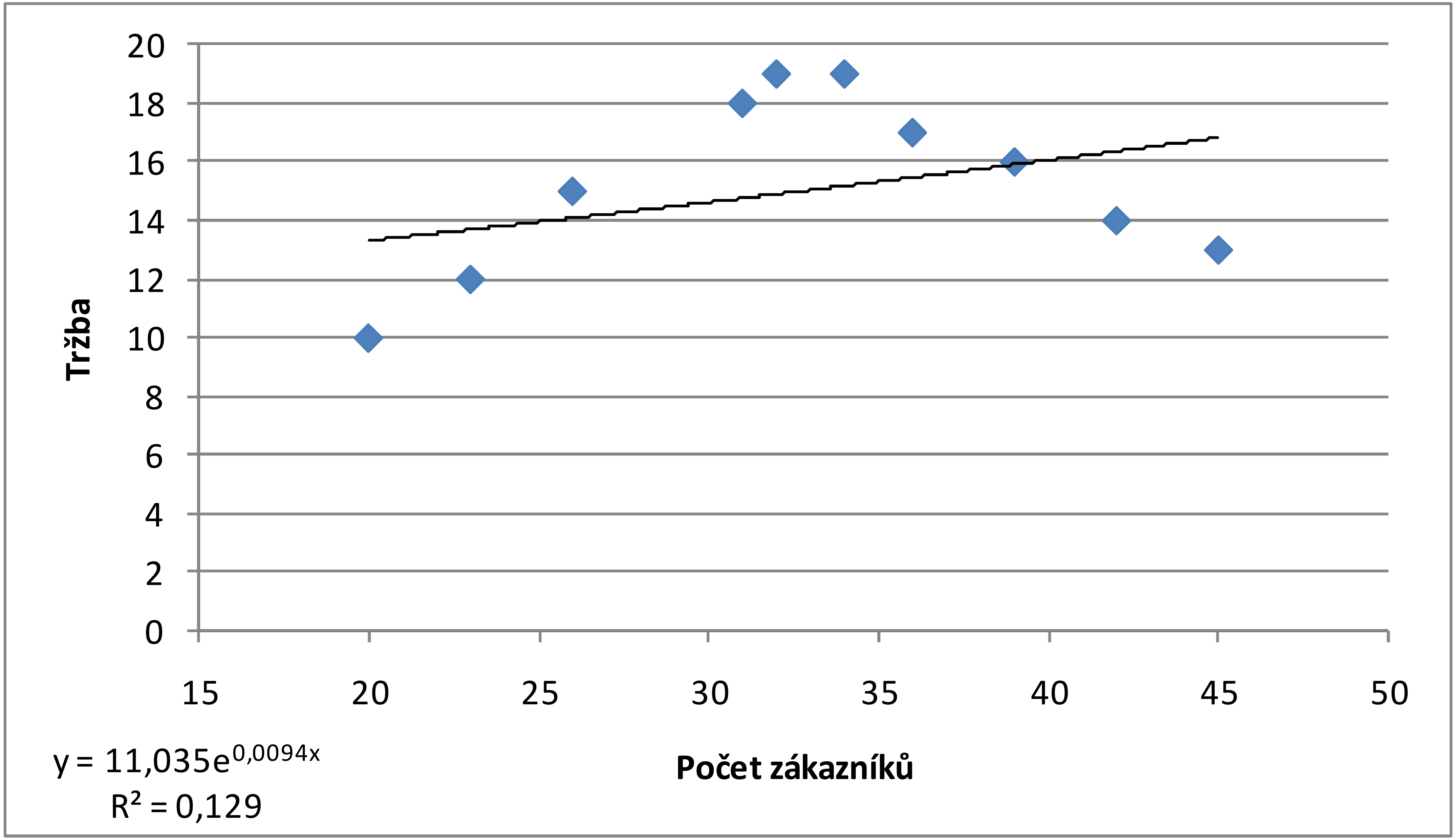
Regresní analýza

- Pro koeficienty regresního modelu platí:

$$\log b_1 = \frac{n\sum_{i=1}^{n} x_i \cdot \log Y_i - \sum_{i=1}^{n} \log Y_i \cdot \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \Rightarrow b_1 = 10^{\left[\frac{n\sum_{i=1}^{n} x_i \cdot \log Y_i - \sum_{i=1}^{n} \log Y_i \cdot \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\right]},$$

$$\log b_0 = \frac{\sum_{i=1}^{n} \log Y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i \cdot \log Y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \Rightarrow b_0 = 10^{\left[\frac{\sum_{i=1}^{n} \log Y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i \cdot \log Y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\right]}.$$

Regresní analýza



Regresní analýza

	x_i	Y_i	$\log Y_i$	x_i^2	$x_i \cdot \log Y_i$
	20	10	1,00000	400	20,00000
	23	12	1,07918	529	24,82117
	26	15	1,17609	676	30,57837
	31	18	1,25527	961	38,91345
	32	19	1,27875	1024	40,92012
	34	19	1,27875	1156	43,47762
	36	17	1,23045	1296	44,29616
	39	16	1,20412	1521	46,96068
	42	14	1,14613	1764	48,13738
	45	13	1,11394	2025	50,12745
Součet	328	153	11,76269	11352	388,23240

$\log b_0$	$\log b_1$
1,04	0,00407
b_0	b_1
11,03	1,01
$\ln b_1$	0,00937

$$\widehat{Y}_i = 11{,}03 \cdot 1{,}01^{x_i}$$

Regresní analýza

	x_i	Y_i	$\log Y_i$	$\log \hat{Y}_i$	$[\log \hat{Y}_i - (\log Y)_p]^2$	$[\log Y_i - (\log Y)_p]^2$
	20	10	1,00000	1,12	0,00271	0,03107
	23	12	1,07918	1,14	0,00159	0,00943
	26	15	1,17609	1,15	0,00077	0,00000
	31	18	1,25527	1,17	0,00005	0,00624
	32	19	1,27875	1,17	0,00001	0,01050
	34	19	1,27875	1,18	0,00002	0,01050
	36	17	1,23045	1,19	0,00017	0,00294
	39	16	1,20412	1,20	0,00064	0,00078
	42	14	1,14613	1,21	0,00140	0,00091
	45	13	1,11394	1,23	0,00247	0,00388
Součet	328	153	11,76269		0,00983	0,07625
Průměr	32,80	15,30	1,17627			

R^2
0,13

Regresní analýza

Vyrovnávací křivka	Index determinace
Přímka	0.09
Parabola	0.93
Hyperbola	0.26
Logaritmická křivka	0.17
Exponenciála	0,13

Na základě vypočtených indexů determinace použijeme pro modelování studované závislosti regresní parabolu.

$$\widehat{Y}_i = -33{,}82 + 3{,}08 x_i - 0{,}05 x_i^2$$

Regresní analýza

- Př. 2 Určete, při jakém počtu zákazníků přicházejících do obchodu za hodinu lze očekávat nejvyšší hodinové tržby.
  - Hledáme maximum na zvolené regresní křivce.

$$\frac{d\widehat{Y}_i}{dx_i} = \frac{d}{dx_i}\left(-33{,}82 + 3{,}08x_i - 0{,}05x_i^2\right) = 3{,}08 - 0{,}1x_i = 0.$$

Tedy:

$$3{,}08 = 0{,}1x_i \Rightarrow x_i = 30{,}8 \doteq 31\,\text{zákazníků}.$$