

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

## TÉMA 4

# ODHAD STŘEDNÍ HODNOTY POMOCÍ TRANSFORMACE

### CO BYSTE MĚLI PO PROSTUDOVÁNÍ TOHOTO TÉMATU UMĚT

1. Proč je nutné v některých případech použít odhad střední hodnoty pomocí nelineární transformace
2. Podstata (princip) transformace
3. Obecné vlastnosti transformační funkce
4. Druhy transformací (především logaritmická a Box Coxova), vhodnost jejich použití a vlastnosti (význam parametru  $\lambda$  u Box Coxovy transformace)
5. Praktický výpočet průměru pomocí Box Coxovy transformace ve Statistice a v Excelu

### OSNOVA

1. Teorie stanovení odhadu střední hodnoty pomocí transformace
2. Druhy transformací a jejich vlastnosti
3. Výpočet ve Statistice a v Excelu
4. Interpretace výsledků

### TEORIE TRANSFORMAČNÍCH ODHADŮ

Viz skripta

[http://user.mendelu.cz/drapela/Statisticke\\_metody/teorie%20text%20II.pdf](http://user.mendelu.cz/drapela/Statisticke_metody/teorie%20text%20II.pdf)

teorie text II, str. 29 - 33

Viz prezentace [http://user.mendelu.cz/drapela/Statisticke\\_metody/Prezentace/zakladni/](http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/)

zde prezentace **EDA.ppt** od snímku 37 dále

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

### PŘÍKLAD VÝPOČTU

Tento typ odhadu parametru ZS se používá tehdy, pokud je rozdělení dat výrazně nesouměrné a (zpravidla) s výraznými extrémními hodnotami, které ale nemůžeme ze souboru vyřadit.

Postup se skládá ze dvou etap:

- 1) ve Statistice si spočítáme transformační konstantu lambda a transformovaná data
- 2) v Excelu spočítáme retransformovanou střední hodnotu včetně intervalového odhadu

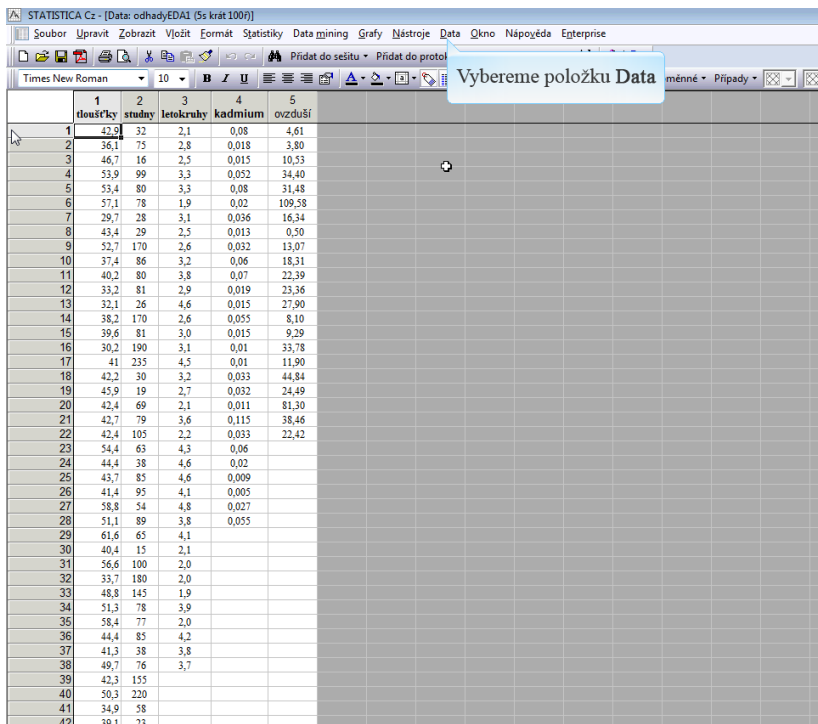
### PŘÍKLAD 1

K výpočtu využijeme proměnnou „Studny“, která je výrazně levostranná a s extrémními hodnotami.

Proměnná „Studny“ je součástí souboru „odhadyEDA1“ zde:

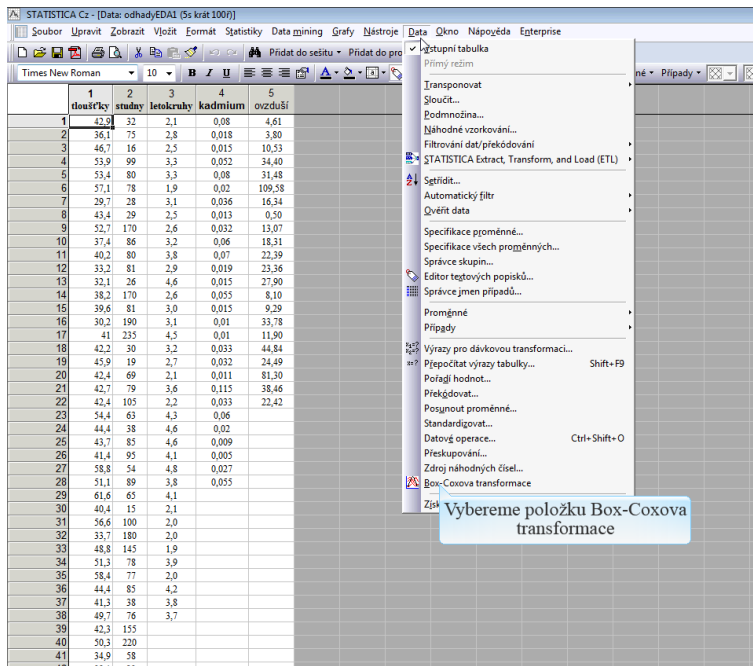
[http://user.mendelu.cz/drapela/Statisticke\\_metody/Data\\_do\\_cviceni/Statistica/odhadyEDA1.sta](http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA1.sta)

Po otevření souboru vybereme položku v hlavním menu „Data“ a dále postupujeme podle obrázků.



	1 tloušťky	2 studny	3 letokruhy	4 kadmium	5 ovzduší
1	42,9	32	2,1	0,08	4,61
2	36,1	75	2,8	0,018	3,80
3	46,7	16	2,5	0,015	10,53
4	53,9	99	3,3	0,052	34,40
5	55,4	80	3,3	0,08	31,48
6	57,1	78	1,9	0,02	109,58
7	29,7	28	3,1	0,036	16,34
8	43,4	29	2,5	0,013	0,50
9	52,7	170	2,6	0,032	13,07
10	37,4	86	3,2	0,06	18,31
11	40,2	80	3,8	0,07	22,39
12	33,2	81	2,9	0,019	23,36
13	32,1	26	4,6	0,015	27,90
14	38,2	170	2,6	0,055	8,10
15	39,6	81	3,0	0,015	9,29
16	30,2	190	3,1	0,01	33,78
17	41	235	4,5	0,01	11,90
18	42,2	30	3,2	0,033	44,84
19	45,9	19	2,7	0,032	24,49
20	42,4	69	2,1	0,011	81,30
21	42,7	79	3,6	0,115	38,46
22	42,4	105	2,2	0,033	22,42
23	54,4	63	4,3	0,06	
24	44,4	38	4,6	0,02	
25	43,7	85	4,6	0,009	
26	41,4	95	4,1	0,005	
27	38,8	54	4,8	0,027	
28	51,1	89	3,8	0,055	
29	61,6	65	4,1		
30	40,4	15	2,1		
31	56,6	100	2,0		
32	33,7	180	2,0		
33	48,8	145	1,9		
34	51,3	78	3,9		
35	58,4	77	2,0		
36	44,4	85	4,2		
37	41,3	38	3,8		
38	49,7	76	3,7		
39	42,3	155			
40	50,3	220			
41	34,9	58			
42	39,1	71			

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



STATISTICA Cz - [Data: odhady\_EDA1 (5x krát1000)]

	1	2	3	4	5
	tloušťky	studny	letokruhy	kadmium	ovzduší
1	42,9	32	2,1	0,08	4,61
2	36,1	75	2,8	0,018	3,80
3	46,7	16	2,5	0,015	10,53
4	53,9	99	3,3	0,052	34,40
5	53,4	80	3,3	0,08	31,48
6	57,1	78	1,9	0,02	109,58
7	29,7	28	3,1	0,036	16,34
8	43,4	29	2,5	0,013	0,50
9	52,7	170	2,6	0,032	13,07
10	37,4	86	3,2	0,06	18,31
11	40,2	80	3,8	0,07	22,39
12	33,2	81	2,9	0,019	23,36
13	32,1	26	4,6	0,015	27,90
14	38,2	170	2,6	0,055	8,10
15	39,6	81	3,0	0,015	9,29
16	30,2	190	3,1	0,01	33,78
17	41	235	4,5	0,01	11,90
18	42,2	30	3,2	0,033	44,84
19	45,9	19	2,7	0,032	24,49
20	42,4	69	2,1	0,011	81,30
21	42,7	79	3,6	0,115	38,46
22	42,4	105	2,2	0,033	22,42
23	54,4	63	4,3	0,06	
24	44,4	38	4,6	0,02	
25	43,7	85	4,6	0,009	
26	41,4	95	4,1	0,005	
27	58,8	54	4,8	0,027	
28	51,1	89	3,8	0,055	
29	61,6	65	4,1		
30	40,4	15	2,1		
31	56,6	100	2,0		
32	33,7	180	2,0		
33	48,8	145	1,9		
34	51,3	78	3,9		
35	58,4	77	2,0		
36	44,4	85	4,2		
37	41,3	38	3,8		
38	49,7	76	3,7		
39	42,3	155			
40	50,3	220			
41	34,9	58			
42	39,1	23			
43	34,3	76			
44	42,5	35			
45	44,8	74			
46	45,2	98			
47	42,7	600			

	1	2	3	4	5
	tloušťky	studny	letokruhy	kadmium	ovzduší
1	42,9	32	2,1	0,08	4,61
2	36,1	75	2,8	0,018	3,80
3	46,7	16	2,5	0,015	10,53
4	53,9	99	3,3	0,052	34,40
5	53,4	80	3,3	0,08	31,48
6	57,1	78	1,9	0,02	109,58
7	29,7	28	3,1	0,036	16,34
8	43,4	29	2,5	0,013	0,50
9	52,7	170	2,6	0,032	13,07
10	37,4	86	3,2	0,06	18,31
11	40,2	80	3,8	0,07	22,39
12	33,2	81	2,9	0,019	23,36
13	32,1	26	4,6	0,015	27,90
14	38,2	170	2,6	0,055	8,10
15	39,6	81	3,0	0,015	9,29
16	30,2	190	3,1	0,01	33,78
17	41	235	4,5	0,01	11,90
18	42,2	30	3,2	0,033	44,84
19	45,9	19	2,7	0,032	24,49
20	42,4	69	2,1	0,011	81,30
21	42,7	79	3,6	0,115	38,46
22	42,4	105	2,2	0,033	22,42
23	54,4	63	4,3	0,06	
24	44,4	38	4,6	0,02	
25	43,7	85	4,6	0,009	
26	41,4	95	4,1	0,005	
27	58,8	54	4,8	0,027	
28	51,1	89	3,8	0,055	
29	61,6	65	4,1		
30	40,4	15	2,1		
31	56,6	100	2,0		
32	33,7	180	2,0		
33	48,8	145	1,9		
34	51,3	78	3,9		
35	58,4	77	2,0		
36	44,4	85	4,2		
37	41,3	38	3,8		
38	49,7	76	3,7		
39	42,3	155			
40	50,3	220			
41	34,9	58			
42	39,1	23			
43	34,3	76			
44	42,5	35			
45	44,8	74			
46	45,2	98			
47	42,7	600			

Dále jsou zde uvedeny některé parametry tranformace:

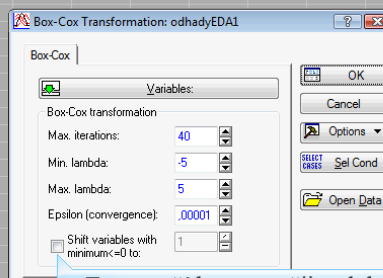
**Max. iterations** - max. počet iterací (opakovaných výpočtů, které se postupně blíží optimálnímu výsledku)

**Min. lambda, max. lambda** - rozsah hodnot lambda, mezi kterými se hledá optimální hodnota  $\lambda$  (tento rozsah s rezervou stačí, většinou se optimální hodnota pohybuje v intervalu -3 až +3)

**Epsilon** - konvergenční kritérium - pokud je mezi výsledky dvou za sebou následujících iterací menší rozdíl než je tato hodnota, výpočet se zastaví jako optimální.

Tyto hodnoty v běžných případech nemusíme měnit.

Objeví se dialogové okno “Box-Cox transformation” a zde vybereme proměnnou určenou k transformaci (“Variables”)



Box-Cox Transformation: odhady\_EDA1

Box-Cox

Variables:

Box-Cox transformation

Max. iterations: 40

Min. lambda: -5

Max. lambda: 5

Epsilon (convergence): .00001

Shift variables with minimum<=0 to:

1

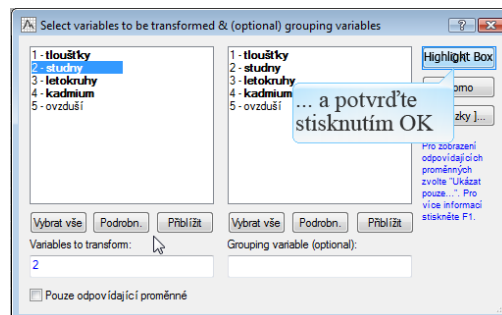
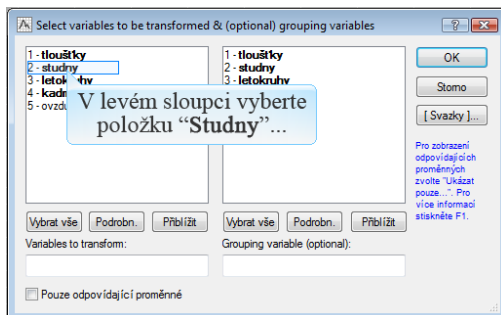
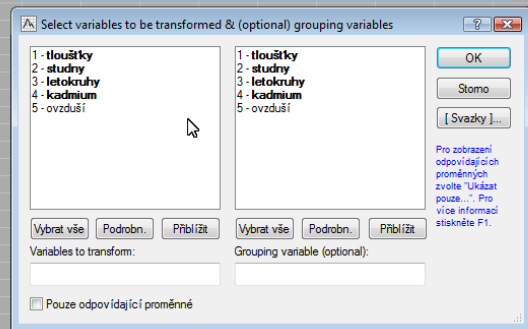
OK Cancel Options

Toto zatřítko se použije tehdy, pokud v datech jsou záporné hodnoty (Box-Coxova transformace pro ně nefunguje). V případě potřeby nastavíme takovou hodnotu posunu, aby všechny hodnoty byly kladné. Vzhledem k tomu, že ke všem hodnotám se připočítá stejná konstanta, vztahy mezi hodnotami zůstanou stejné

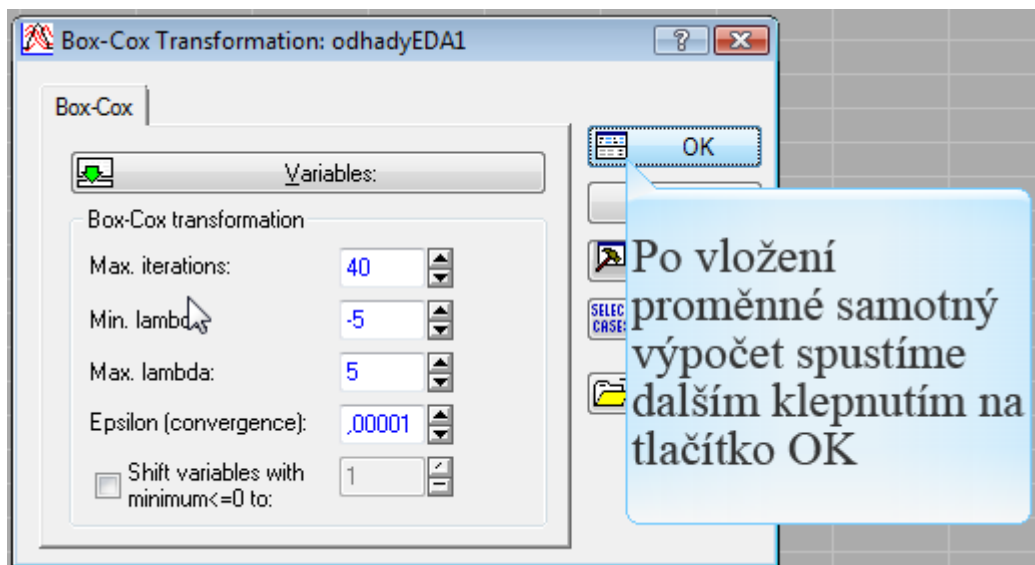
## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

	1 tloušťky	2 studny	3 letokruhy	4 kadmium	5 ovzduší
1	42,9	32	2,1	0,08	4,61
2	36,1	75	2,8	0,018	3,80
3	46,7	16	2,5	0,015	10,53
4	53,9	99	3,3	0,052	34,40
5	53,4	80	3,3	0,08	31,48
6	57,1	78	1,9	0,02	109,58
7	29,7	28	3,1	0,036	16,34
8	43,4	29	2,5	0,013	0,50
9	52,7	170	2,6	0,032	13,07
10	37,4	86	3,2	0,06	18,31
11	40,2	80	3,8	0,07	22,39
12	33,2	81	2,9	0,019	23,36
13	32,1	26	4,6	0,015	27,90
14	38,2	170	2,6	0,055	8,10
15	39,6	81	3,0	0,015	9,29
16	30,2	190	3,1	0,01	33,78
17	41	235	4,5	0,01	11,90
18	42,2	30	3,2	0,033	44,84
19	45,9	19	2,7	0,032	24,49
20	42,4	69	2,1	0,011	81,30
21	42,7	79	3,6	0,115	38,46
22	42,4	105	2,2	0,033	22,42
23	54,4	63	4,3	0,06	
24	44,4	38	4,6	0,02	
25	43,7	85	4,6	0,009	
26	41,4	95	4,1	0,005	
27	58,8	54	4,8	0,027	
28	51,1	89	3,8	0,055	
29	61,6	65	4,1		
30	40,4	15	2,1		
31	56,6	100	2,0		
32	33,7	180	2,0		
33	48,8	145	1,9		
34	51,3	78	3,9		
35	58,4	77	2,0		
36	44,4	85	4,2		
37	41,3	38	3,8		
38	49,7	76	3,7		
39	42,3	155			
40	50,3	220			

Otevřete se okno "The Select variables to be transformed (optional) grouping variables"  
(Vyberte proměnné k transformaci a (volitelně) třídící proměnné).  
Vzhledem k tomu, že žádné třídění nepotřebujeme provádět, vybereme pouze proměnnou pro transformaci v levém sloupci.

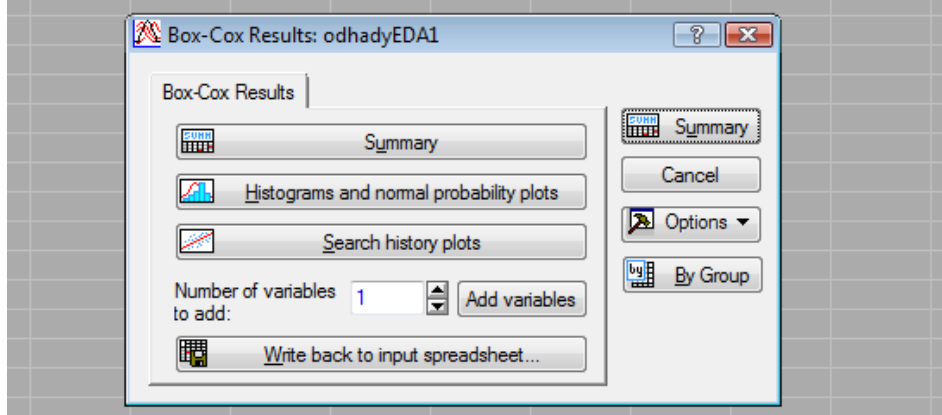


## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



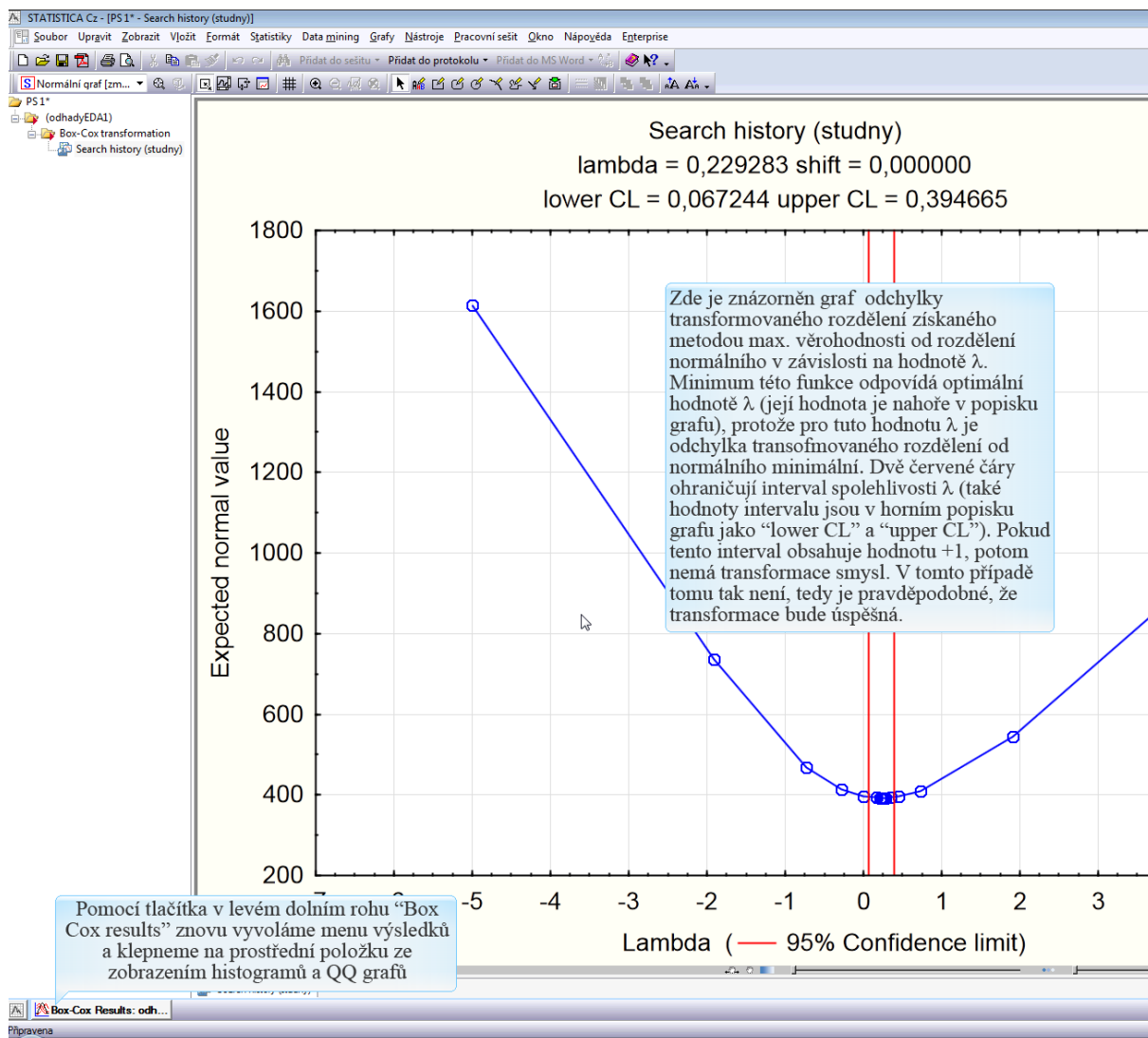
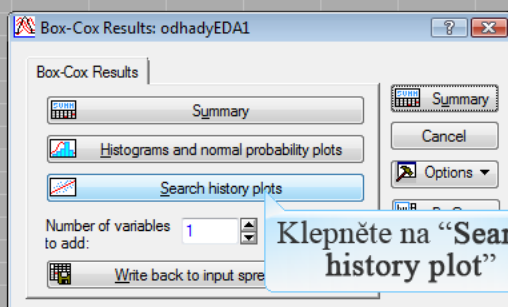
Objeví se okno s výsledky transformace:

- 1) Summary - tabulka výsledků (hodnota  $\lambda$  a její interval spolehlivosti)
- 2) Histograms and normal probability plots (Histogramy a QQ grafy pro normální rozdělení) - zde jsou porovnány tyto grafy pro původní a transformovaná data
- 3) Search history plots (graf hledání řešení optimální hodnoty  $\lambda$  včetně vykreslení jejího IS)



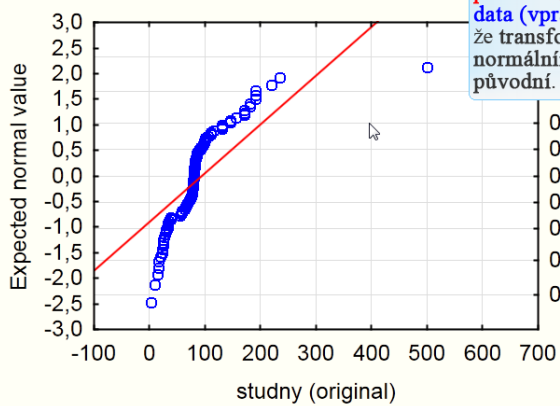
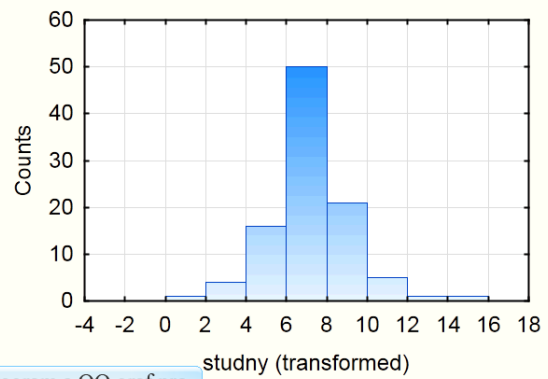
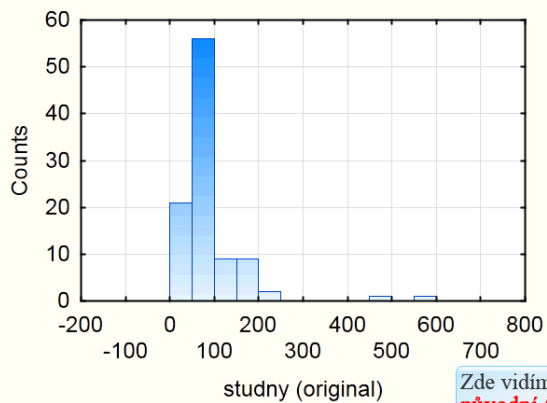
## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nyní si prohlédneme výsledky:

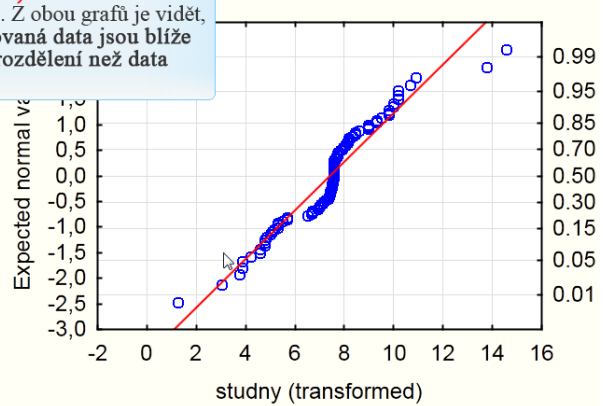


## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Histogram & normal probability plots (studny)  
lambda = 0,229283 shift = 0,000000

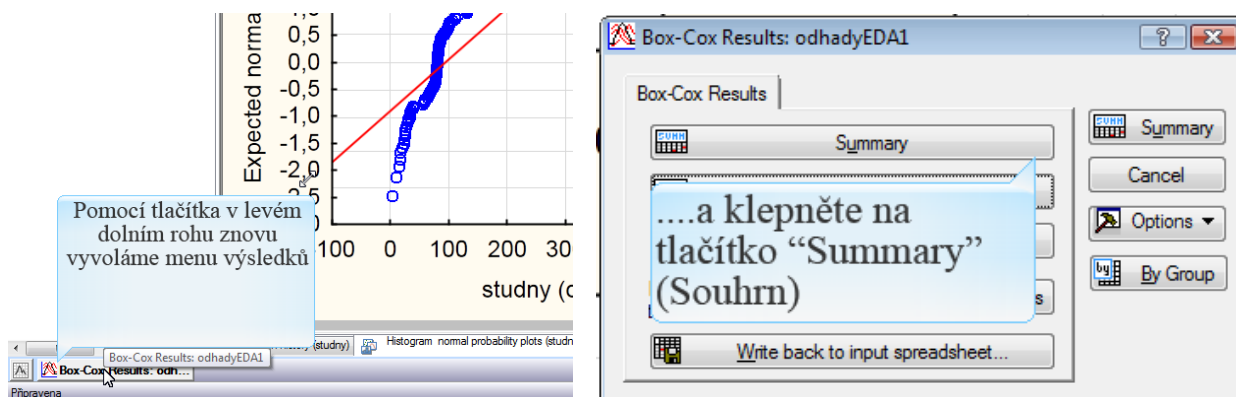


Zde vidíme histogram a QQ graf pro **původní (vlevo)** a **transformovaná data (vpravo)**. Z obou grafů je vidět, že transformovaná data jsou blíže normálnímu rozdělení než data původní.



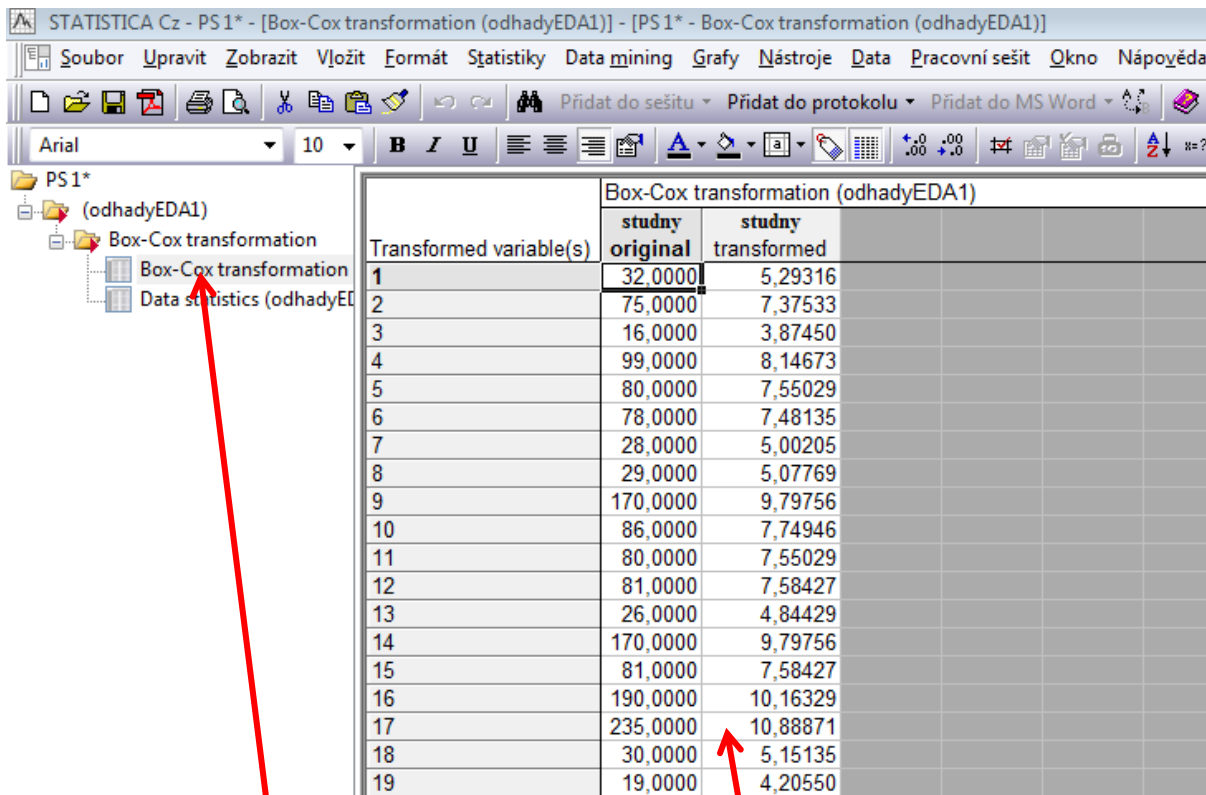


## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Tento výsledek (a hodnoty transformovaných dat na dalším listu nazvaném “Box-Cox transformation”) jsou konečné výsledky, které nám poskytne STATISTICA.

Retransformaci střední hodnoty a výpočet jejího intervalového odhadu musíme spočítat v Excelu. Pro tyto účely musíme do Excelu zkopírovat následující:  
hodnotu lambda  
oba sloupce (původní i transformovaná data) z listu “Box-Cox transformation”



Transformed variable(s)	Box-Cox transformation (odhadyEDA1)	
	studny original	studny transformed
1	32,0000	5,29316
2	75,0000	7,37533
3	16,0000	3,87450
4	99,0000	8,14673
5	80,0000	7,55029
6	78,0000	7,48135
7	28,0000	5,00205
8	29,0000	5,07769
9	170,0000	9,79756
10	86,0000	7,74946
11	80,0000	7,55029
12	81,0000	7,58427
13	26,0000	4,84429
14	170,0000	9,79756
15	81,0000	7,58427
16	190,0000	10,16329
17	235,0000	10,88871
18	30,0000	5,15135
19	19,0000	4,20550

Hodnotu  $\lambda$  z listu Data Statistics a hodnoty transformovaných dat na dalším listu nazvaném “Box-Cox transformation”) jsou konečné výsledky, které nám poskytne STATISTICA.



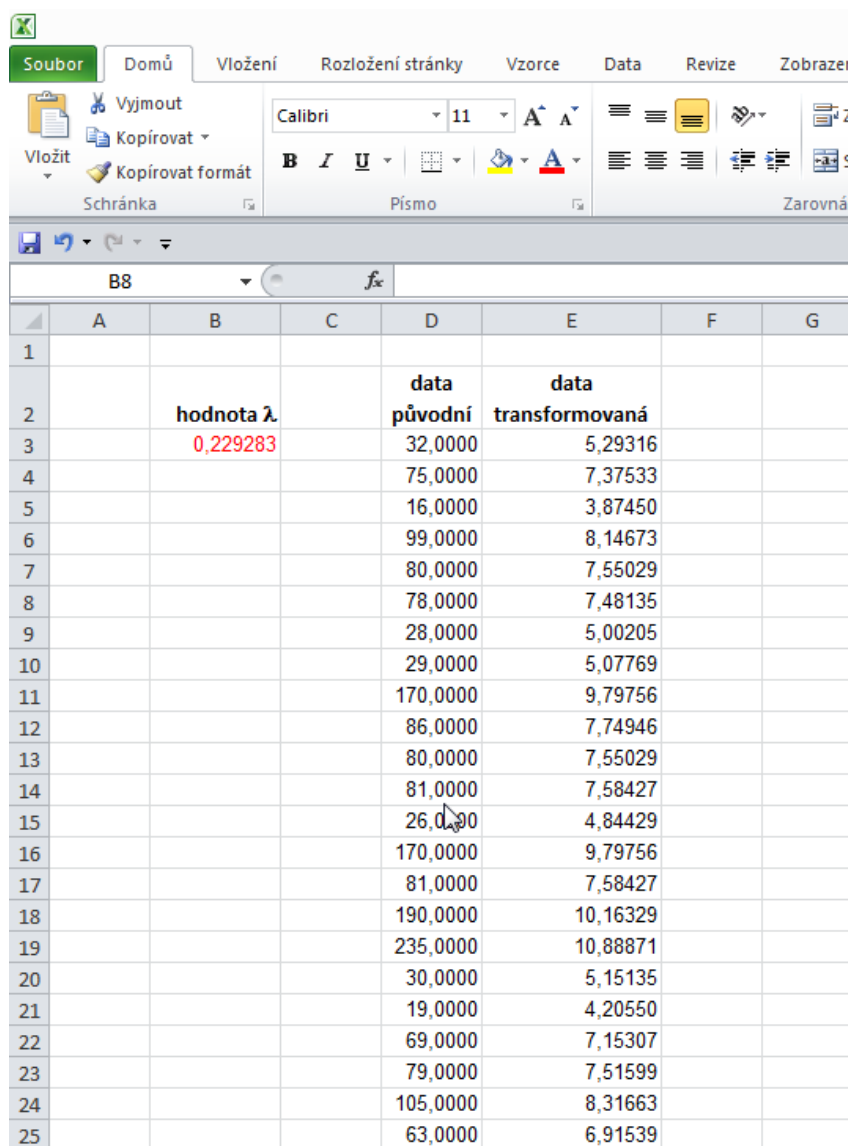
## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

**Retransformaci střední hodnoty a výpočet jejího intervalového odhadu musíme spočítat v Excelu.**

Pro tyto účely musíme do Excelu zkopírovat následující:

- hodnotu  $\lambda$
- oba sloupce (původní i transformovaná data) z listu "Box-Cox transformation"

Následující obrázek ukazuje potřebná data vložená do Excelu (s malou úpravou formátu – např. úpravu nadpisů, barva písma apod.)

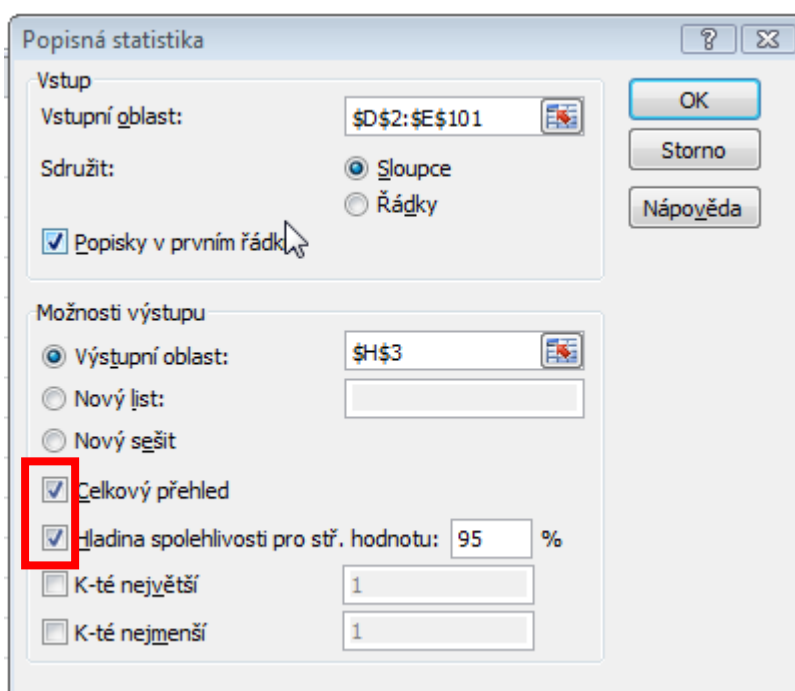


	A	B	C	D	E	F	G
1							
2		<b>hodnota <math>\lambda</math></b>		<b>data původní</b>	<b>data transformovaná</b>		
3		0,229283		32,0000	5,29316		
4				75,0000	7,37533		
5				16,0000	3,87450		
6				99,0000	8,14673		
7				80,0000	7,55029		
8				78,0000	7,48135		
9				28,0000	5,00205		
10				29,0000	5,07769		
11				170,0000	9,79756		
12				86,0000	7,74946		
13				80,0000	7,55029		
14				81,0000	7,58427		
15				26,0000	4,84429		
16				170,0000	9,79756		
17				81,0000	7,58427		
18				190,0000	10,16329		
19				235,0000	10,88871		
20				30,0000	5,15135		
21				19,0000	4,20550		
22				69,0000	7,15307		
23				79,0000	7,51599		
24				105,0000	8,31663		
25				63,0000	6,91539		

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Pro oba datové soubory spočítáme běžné popisné charakteristiky pomocí „Analýzy dat“. Nezapomeňte zatrhnout i „Hladina spolehlivosti střední hodnoty“, aby se spočítala polovina intervalu spolehlivosti střední hodnoty.

Dialogové okno analýzy bude vypadat podobně jako toto (pouze Vstupní oblast a Výstupní oblast budou vyplněny podle vaší skutečnosti, tj. bude zde váš rozsah buněk pro analyzovaná data (Vstupní oblast) a zadán levý horní roh výstupní oblasti (Výstupní oblast):



Popisná statistika

Vstup

Vstupní oblast:

Sdružit:  Sloupce  Řádky

Popisky v prvním řádku

Možnosti výstupu

Výstupní oblast:

Nový list:

Nový sešit

Celkový přehled

Hladina spolehlivosti pro stř. hodnotu:  %

K-té největší:

K-té nejmenší:

OK Storno Nápořádá

**Pozn. – Při praktické analýze bychom vše počítali pouze pro transformovaná data, zde pro účely srovnání výsledků budeme vše počítat i pro původní data.**

Po určité úpravě formátu (odmazání duplicitních popisků), bude výstup vypadat takto:

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

	data původní	data transformovaná
Stř. hodnota	93,05051	7,419192135
Chyba stř. hodnoty	8,192695	0,199658335
Medián	80	7,550292162
Modus	80	7,550292162
Směr. odchylka	81,51629	1,986575352
Rozptyl výběru	6644,906	3,94648163
Špičatost	20,29007	2,345911827
Šikmost	3,869086	0,230096484
#ODKAZ!	597	13,2957008
Minimum	3	1,249366605
Maximum	600	14,54506741
Součet	9212	734,5000213
Počet	99	99
Hladina spolehlivosti (95,0%)	16,25814	0,396215468

Důležité jsou pro nás hlavně první a poslední řádek, z nich vypočítáme intervalový odhad pro původní i transformovaná data.

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

	data původní	data transformovaná
Stř. hodnota	93,05051	7,419192135
Chyba stř. hodnoty	8,192695	0,199658335
Medián	80	7,550292162
Modus	80	7,550292162
Směr. odchylka	81,51629	1,986575352
Rozptyl výběru	6644,906	3,94648163
Špičatost	20,29007	2,345911827
Šikmost	3,869086	0,230096484
#ODKAZ!	597	13,2957008
Minimum	3	1,249366605
Maximum	600	14,54506741
Součet	9212	734,5000213
Počet	99	99
Hladina spolehlivosti (95,0%)	16,25814	0,396215468
dolní hranice IS	76,79237	7,022976667
horní hranice IS	109,3086	7,815407603

**Výsledek pro původní data je již konečný** (červená čísla). Znamená to, že pokud bychom ignorovali nesouměrnost dat a extrémní hodnoty, získali bychom výsledek, že střední hodnota se rovná 93,1 mg/l s 95-ti % intervalem spolehlivosti (IS) (76,8;109,3) mg/l.

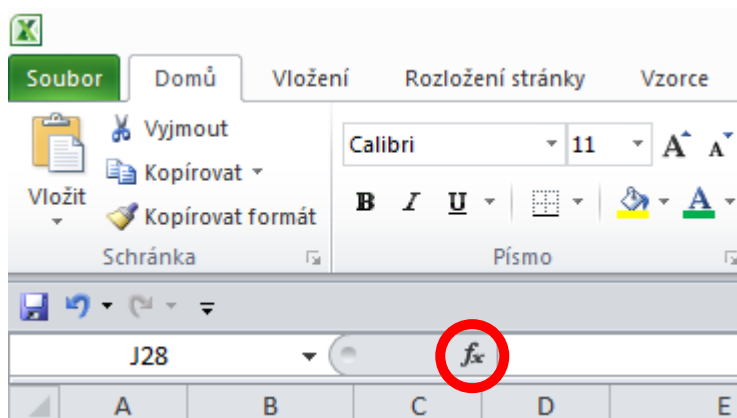
**Zelená čísla představují výsledek pro transformovaná data.** Všimněte si, že transformovaná data mají opravdu daleko vhodnější rozdělení (koef. šikmosti je 0,23 oproti původní hodnotě 3,87, také koef. špičatosti se podstatně snížil z 20,3 na 2,3, takže rozdělení transformovaných dat je podstatně blíže normálnímu rozdělení). Nicméně, střední hodnota a její IS je vypočítán pouze pro transformovaná data, která nemají s původními daty „nic“ společného (jsou samozřejmě spolu „svázané“ transformační funkcí). **Proto je nutné tyto hodnoty retransformovat do původních hodnot.**

Při retransformaci vycházíme z původního vzorce Box-Coxovy transformace, z níž vyjádříme vzorec pro výpočet původní hodnoty  $x$  ( $g$  je transformovaná hodnota, hodnotu  $\lambda$  máme vypočítanou z programu STATISTICA):

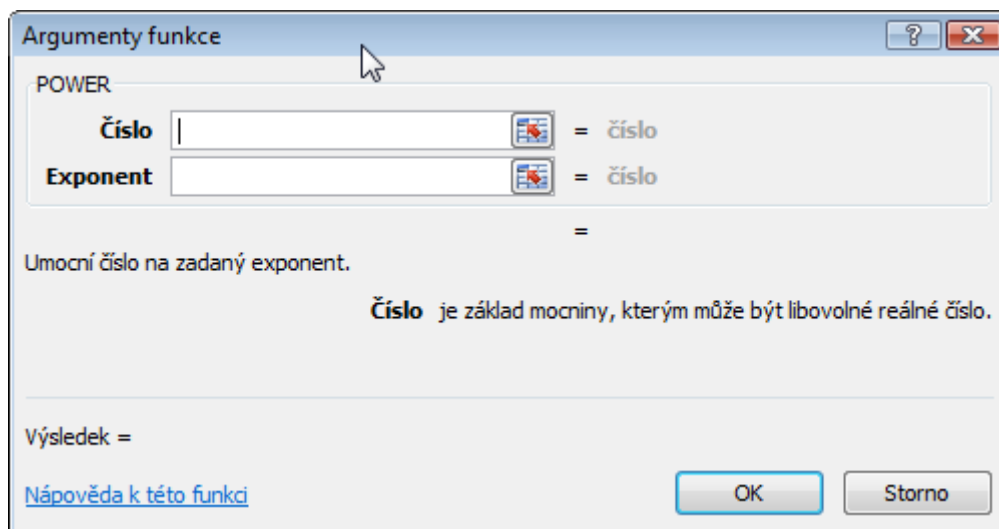
$$g = \frac{x^\lambda - 1}{\lambda} \Rightarrow x = \sqrt[\lambda]{g \cdot \lambda + 1}$$

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Jediným problémem bude výpočet „lambda odmocniny“. Známa funkce Excelu ODMOCNINA() je pouze druhá odmocnina, takže musíme použít malý „trik“, kdy použijeme místo odmocniny umocnění na převrácenou hodnotu – tj.  $\sqrt[y]{x} = x^{\frac{1}{y}}$ . K tomu využijeme funkci Excelu POWER. Funkci vložíme pomocí tlačítka pro vložení funkce



Dále nastavíme „Vybrat kategorií“ na „Matematické“ a objeví se následující okno



Do okna „Číslo“ se vloží číslo pod odmocninou, do okna „Exponent“ se vloží hodnota  $1/\lambda$ .

Následující obrázek ukazuje vzorec pro výpočet retransformovaného průměru. Vycházíme z odkazu na transformovaný průměr (zde J5) a na lambdu (zde B3). „Dolary“ v lambdy jsou proto, že i při kopírování vzorce půjde odkaz vždy na tuto buňku (absolutní odkaz). „Dolary“ vložíme do příslušného odkazu tak, že umístíme kurzor kdekoliv v odkazu a zmáčkneme funkční klávesu F4.

Stejným způsobem vypočítáme retransformované hranice IS, jen hodnotu průměru (J3) zaměníme za „zelené“ hodnoty dolní a horní hranice (zde J20 a J21).

Postup ukazují dva následující obrázky.

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

transformace - Microsoft Excel

POWER =POWER(J5\*\$B\$3+1;1/\$B\$3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
		hodnota λ		data původní	data transformovaná				data původní	data transformovaná			
2		0,229283		32,0000	5,29316								
3				75,0000	7,37533								
4				16,0000	3,87450			Stř. hodnota	93,05051	7,419192135			
5				99,0000	8,14673			Chyba stř. hodnoty	8,192695	0,199658335			
6				80,0000	7,55029			Medián	80	7,550292162			
7				78,0000	7,48135			Modus	80	7,550292162			
8				28,0000	5,00205			Směr. odchylka	81,51629	1,986575352			
9				29,0000	5,07769			Rozptyl výběru	6644,906	3,94648163			
10				170,0000	9,79756			Špičatost	20,29007	2,345911827			
11				86,0000	7,74946			Šikmost	3,869086	0,230096484			
12				80,0000	7,55029			#ODKAZ!	597	13,2957008			
13				81,0000	7,58427			Minimum	3	1,249366605			
14				26,0000	4,84429			Maximum	600	14,54506741			
15				170,0000	9,79756			Součet	9212	734,5000213			
16				81,0000	7,58427			Počet	99	99			
17				190,0000	10,16329			Hladina spolehlivosti (95,0%)	16,25814	0,396215468			
18				235,0000	10,88871								
19				30,0000	5,15135			dolní hranice IS	76,79237	7,022976667			
20				19,0000	4,20550			horní hranice IS	109,3086	7,815407603			
21				69,0000	7,15307								
22				79,0000	7,51599			střední hodnota retransformovaná		=POWER(J5*\$B\$3+1;1/\$B\$3)			
23				105,0000	8,31663			dolní hranice IS retransformovaný					
24				63,0000	6,91539			horní hranice IS retransformovaný					
25				38,0000	5,68117								
26				80,0000	7,55029								

	hodnota λ	data původní	data transformovaná			data původní	data transformovaná
	0,229283	32,0000	5,29316				
		75,0000	7,37533				
		16,0000	3,87450	Stř. hodnota		93,05051	7,419192135
		99,0000	8,14673	Chyba stř. hodnoty		8,192695	0,199658335
		80,0000	7,55029	Medián		80	7,550292162
		78,0000	7,48135	Modus		80	7,550292162
		28,0000	5,00205	Směr. odchylka		81,51629	1,986575352
		29,0000	5,07769	Rozptyl výběru		6644,906	3,94648163
		170,0000	9,79756	Špičatost		20,29007	2,345911827
		86,0000	7,74946	Šikmost		3,869086	0,230096484
		80,0000	7,55029	#ODKAZ!		597	13,2957008
		81,0000	7,58427	Minimum		3	1,249366605
		26,0000	4,84429	Maximum		600	14,54506741
		170,0000	9,79756	Součet		9212	734,5000213
		81,0000	7,58427	Počet		99	99
		190,0000	10,16329	Hladina spolehlivosti (95,0%)		16,25814	0,396215468
		235,0000	10,88871				
		30,0000	5,15135	dolní hranice IS		76,79237	7,022976667
		19,0000	4,20550	horní hranice IS		109,3086	7,815407603
		69,0000	7,15307				
		79,0000	7,51599	střední hodnota retransformovaná			76,23028
		105,0000	8,31663	dolní hranice IS retransformovaný			65,66383
		63,0000	6,91539	horní hranice IS retransformovaný			88,06129





## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Výsledkem je, že hodnota transformovaného průměru je 76,2 mg/l s IS (65,7;88,1) mg/l.

Všimněte si, že výsledný průměr a jeho IS je posunut výrazně doleva a je užší (IS pro původní data má „šířku“ asi 32,5, transformovaný IS má šířku 22,4. Transformovaný průměr se výrazně přiblížil mediánu souboru (hodnotě, která je „robustní“ vůči extrémním hodnotám) 76,3 oproti 80 (medián).

Další výraznou vlastností transformovaného IS je fakt, že je nesouměrný vůči výběrovému průměru (čímž „ctí“ nesouměrnost původních dat, která jsou silně levostranná) – dolní část IS má šířku asi 10,5, horní část asi 11,9. Je to logické, neboť v levé části souboru jsou data více koncentrována, takže jsou soustředěna na menší části číselné osy.

### PŘÍKLAD 2

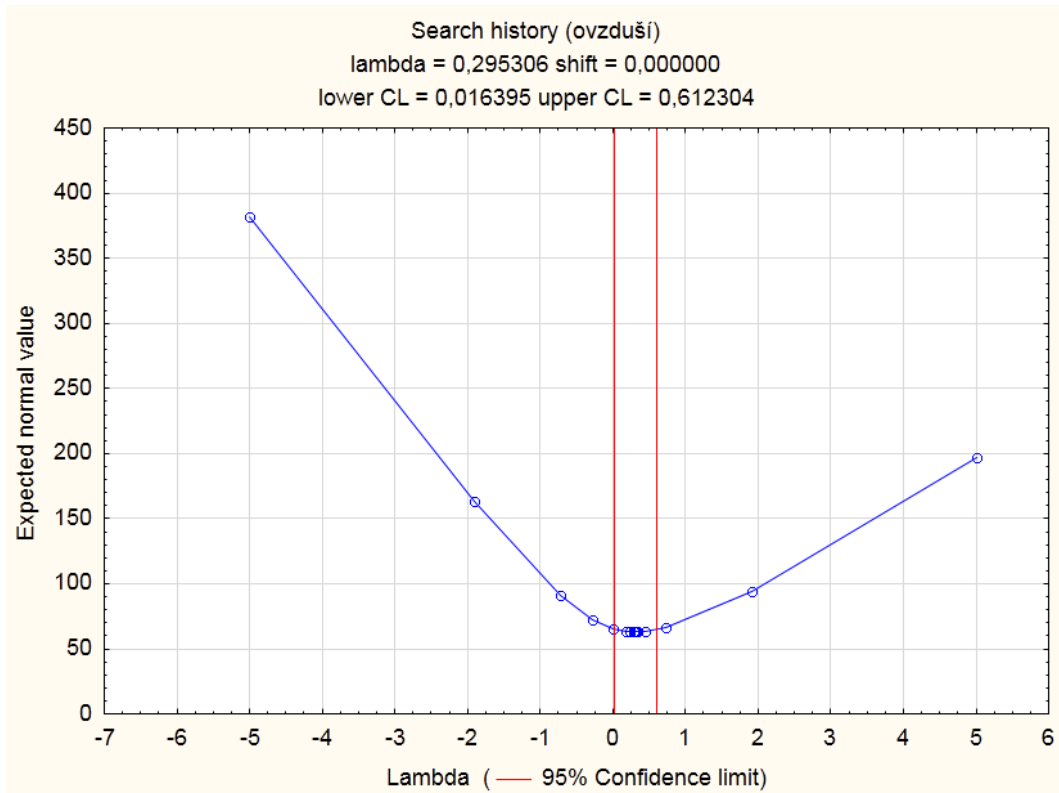
K výpočtu použijeme proměnnou „Ovzduší“ ze stejného souboru „odhadyEDA1“ jako v příkladu 1:

[http://user.mendelu.cz/drapela/Statisticke\\_metody/Data\\_do\\_cviceni/Statistica/odhadyEDA1.sta](http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA1.sta)

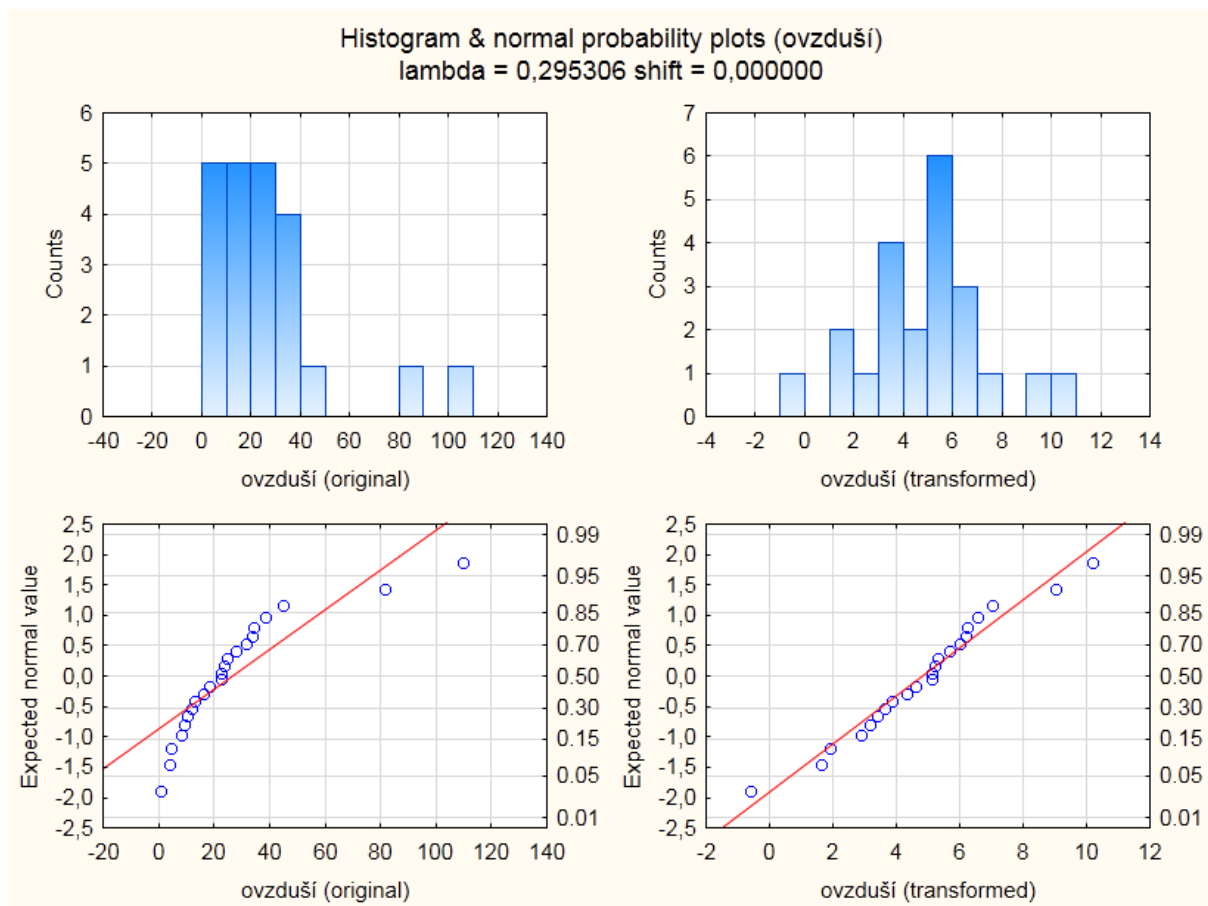
Postup výpočtu již nebude podrobně komentován, pouze výsledky. **Celý výpočet si podle návodu v příkladu 1 zkuste sami.** Vzhledem k tomu, že tato proměnná obsahuje relativně málo dat, které se dají poměrně přehledně graficky zobrazit, bude srovnání výsledků prezentováno i powerpointové prezentaci. U předchozího souboru toto není vzhledem k množství dat a přílišné „extrémnosti“ odlehklých hodnot z důvodů grafické přehlednosti možné.

Stanovení hodnoty  $\lambda$ :

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



Na základě QQ grafu a histogramu vidíme, že proměnná „ovzduší“ je silně levostranná se dvěma zřetelnými odlehlými hodnotami. Vysoké jsou hodnoty jak koef. špičatosti, tak šikmosti. Proto není možné použít „klasický“ souměrný interval spolehlivosti střední hodnoty (který předpokládá normální rozdělení).

## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Charakteristika	Původní data "ovzduší"	Transformovaná data "ovzduší"
Stř. hodnota	26,86	4,84
Chyba stř. hodnoty	5,44	0,51
Medián	22,41	5,10
Směr. odchylka	25,54	2,39
Rozptyl výběru	652,17	5,71
Špičatost	5,04	0,90
Šikmost	2,11	0,04
Variační rozpětí	109,08	10,79
Minimum	0,50	- 0,63
Maximum	109,58	10,17
Součet	590,85	106,41
Počet	22,00	22,00
Hladina spolehlivosti (95,0%)	11,32	1,06

Grafické výstupy ukazují, že optimální hodnota  $\lambda$  je 0,295 s intervalem spolehlivosti (0,016;0,612). Z toho vyplývá, že transformace může být úspěšná (IS  $\lambda$  neobsahuje hodnotu +1). V Excelu spočítáme retransformované hodnoty průměru a jeho intervalový odhad.

Výše uvedená tabulka dokládá, že transformací se prakticky odstranilo sešikmení dat (koef. šikmosti klesl z 2,11 na 0,04) a výrazně se snížila špičatost (z 5,04 na 0,90). Proto pro transformovaná data použijeme "klasický" výpočet IS a výsledky retransformujeme.

Odhad	Původní data "ovzduší"	Transformovaná data "ovzduší"
střední hodnota	<b>26,86</b>	4,84
dolní hranice IS stř. hodnoty	<b>15,53</b>	3,78
horní hranice IS stř. hodnoty	<b>38,18</b>	5,90
střední hodnota retransformovaná		20,17
dolní hranice IS stř. hodnoty retransformovaná		12,64
horní hranice IS stř. hodnoty retransformovaná		30,41

Grafické porovnání IS pro původní a transformovaná data je v prezentaci

[http://user.mendelu.cz/drapela/Statisticke\\_metody/Prezentace/zakladni/BoxCox\\_transformace\\_ovzdusi.ppsx](http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/BoxCox_transformace_ovzdusi.ppsx)

Tato prezentace je již připravena k přehrávání, stačí ji spustit a poté, co si příslušný obrázek prohlédnete a přečtete doprovodný text, postupovat dále mačkáním klávesy Enter).

Pokud máte starší verzi PowerPointu než 2007, prezentace nepůjde přehrát. Další možnosti, jak přehrát prezentaci bez PowerPointu je zde:



## INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

<https://support.microsoft.com/cs-cz/office/zobrazen%C3%AD-prezentace-bez-powerpointu-2f1077ab-9a4e-41ba-9f75-d55bd9b231a6>