



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 λ u Box Coxovy transfomace)
- 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

teorie text II, str. 29 - 33

Viz prezentace <u>http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/</u> zde prezentace **EDA.ppt od snímku 37 dále**



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:

 $\underline{http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA1.st}{\underline{a}}$

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

STATISTICA Cz - [Data: odhadyEDA1 (5s krát 1007)]																
Soubor Upravit Zobrazit Vložit Formát Statistiky Datamining Grafy Nástroje Data Okno Nápovéda Enterprise																
D 🖻 📮	🗅 😂 🔚 🔁 🙆 这 🕺 🖻 🔍 🍼 🖘 🏘 Přídat do sešitu 🝷 Přídat do proto															
Times New	Roman	-	10 - B	сп		A 194	ð - 🖬	8	Vyber	eme no	oložku	Data	měnné 🔻	Přínady -	RX -	
							<u> </u>			enne p	JIOLINA	Duiu		mpouy	1XXX	
	1 tloušťky	2 studny	3 letokruhy	4 kadmium	5 ovzduší											
N 1	42,9	32	2,1	0,08	4,61											
13 2	36,1	75	2,8	0,018	3,80											
3	46,7	16	2,5	0,015	10,53			<u> </u>								
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	23,30											
13	32,1	26	4,6	0,015	27,90											
14	38,2	1/0	2,0	0,055	8,10											
15	39,6	81	3,0	0,015	9,29											
10	30,2	190	3,1	0,01	33,/8											
17	42.2	235	4,5	0,01	11,90											
10	42,2	30	3,2	0,033	44,84											
20	42,5	60	2,7	0,032	£4,47 81.20											
20	42,4	70	3.6	0,011	38.46											
21	42,7	105	2.2	0,022	22,42											
23	54.4	63	4.3	0.06	22,42											
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														



A STATISTICA CT.





mes New	Roman	-	10 👻 🖪	IU		₫ 4-	🖄 r 💽 r 🚫		né • Případy • 🔯 🔤
	1	2	3	4	5				<u>T</u> ransponovat
	tloušť ky	studny	letokruhy	kadmium	ovzduší				Sloučit
1	42.0	32	21	0.08	4.61				Podmnožina
2	36.1	75	2.8	0.018	3.80				Náhodné vzorkování
3	46.7	16	2.5	0.015	10.53				Filtrování dat/překódování •
4	53.9	99	3.3	0.052	34.40				STATISTICA Extract, Transform, and Load (ETL)
5	53,4	80	3,3	0,08	31,48			A 1	C
6	57.1	78	1.9	0.02	109,58			Z+	Sgtridit
7	29,7	28	3,1	0,036	16,34				Automatický filtr
8	43,4	29	2,5	0,013	0,50				Qvěřít data •
9	52.7	170	2.6	0.032	13.07				Specifikace proménné
10	37,4	86	3,2	0,06	18,31				Specifikace všech proměnných
11	40,2	80	3,8	0,07	22,39				opecinkace vieco promennych
12	33,2	81	2,9	0,019	23,36			0	spravce skupin
13	32,1	26	4,6	0,015	27,90			\geq	Editor testových popisků
14	38,2	170	2,6	0,055	8,10				Správce jmen případů
15	39,6	81	3,0	0,015	9,29				Proměnné
16	30,2	190	3,1	0,01	33,78				Pönady
17	41	235	4,5	0,01	11,90				- inpguy
18	42,2	30	3,2	0,033	44,84			24=?	Výrazy pro dávkovou transformaci
19	45,9	19	2,7	0,032	24,49			8:?	Přepočítat výrazy tabulky Shift+F9
20	42,4	69	2,1	0,011	81,30				Pořadí hodnot
21	42,7	79	3,6	0,115	38,46				Překódovat
22	42,4	105	2,2	0,033	22,42				Posupout proménné
23	54,4	63	4,3	0,06					Chandradian at
24	44,4	38	4,6	0,02					Standardigovat
25	43,7	85	4,6	0,009					Datove operace Ctri+Snift+O
26	41,4	95	4,1	0,005					Preskupovani
27	58,8	54	4,8	0,027					Zdroj náhodných čísel
28	51,1	89	3,8	0,055				200	Box-Coxova transformace
29	61,6	65	4,1						7/st TV 1 D C
30	40,4	15	2,1						Vybereme položku Box-Coxova
31	56,6	100	2,0						transformace
32	33,7	180	2,0						amoreannee
33	48,8	145	1,9						
34	51,3	/8	3,9						
35	58,4	11	2,0						
36	44,4	85	4,2						
3/	41,3	38	3,8						
38	49,7	/0	3,7						
39	42,5	132							
40	30,3	220							
41	54,9	38							

	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	Dále isou zde uvedeny některé	
13	32,1	26	4,6	0,015	27,90	perometry tranfromaco:	
14	38,2	170	2,6	0,055	8,10	parametry trainfollace.	Objeví se dialogové okno "Box-Cox
15	39,6	81	3,0	0,015	9,29		Objevi se dialogove okilo Dox-Cox
10	50,2	225	3,1	0,01	33,78	Max iterations - max, počet	transformation" a zde vybereme promennou
18	42.2	30	3.2	0.033	11,50	itaraaí (anakovaných vymočtů	určenou k transformaci ("Variables")
10	45.0	10	2.7	0.032	24.49	neraci (opakovanych vypociu,	
20	42,5	69	2,7	0.011	81.30	které se postupně blíží	
21	42.7	79	3.6	0 115	38.46	optimálnímu výsledku)	M Box-Cox Transformation: odhadyEDA1
22	42.4	105	2.2	0.033	22.42	optimitation () situatio)	
23	54,4	63	4.3	n_0.06			Box-Cox
24	44,4	38	4,6	0.02		Min. lambda, max. lambda -	
25	43,7	85	4,6	0,009		rozsah hodnot lambda mezi	
26	41,4	95	4,1	0,005		Istariuni as bladá antimální	Box-Cox transformation
27	58,8	54	4,8	0,027		kterynn se meda optimalin	Max. iterations: 40 🍝 🛛 Apptions 🔻
28	51,1	89	3,8	0,055		hodnota λ (tento rozsah s	
29	61,6	65	4,1			rezervou stačí většinou se	
30	40,4	15	2,1			antimální hadnata nahybuja y	Max. lambda: 5 🚔
31	56,6	100	2,0			opulnalni nodnota poliybuje v	Ensilon (convergence): 00001
32	33,7	180	2,0			intervalu -3 až +3)	
33	48,8	145	1,9			· ·	Shift variables with 1
34	51,3	78	3,9			Trailan Ironwanaan Xn/ Irrit/riyan	
30	38,4	//	2,0			Epsilon - konvergenein kriterium	Toto zatržítko se použije tehdy, pokud v
30	44,4	20	4,2			- pokud je mezi výsledky dvou za	1 to 1 in the set pouzife tendy, pokud v
38	41,5	76	2,0			sebou následujících iterací menší	datech jsou zaporne hodnoty (Box-Coxova
30	42,7	155	3,1			negatil než is teta le dueta sejus žet	transformace pro ně nefunguje). V případě
40	50.3	220				rozun nez je tato nodnota, vypočet	notřehy nastavíme takovou hodnotu nosunu
40	34.9	58				se zastaví jako optimální.	
42	39.1	23				J 1	aby všechny hodnoty byly kladné. Vzhledem
43	34.3	76				Testa hada stara hXXadah	k tomu, že ke všem hodnotám se připočítá
44	42.5	35				Tyto nodnoty v beznych	steiné konstante, vztahy mezi hodnotemi
45	44,8	74				případech nemusíme měnit.	stejna konstanta, vztany mezi nodnotanii
46	45,2	98				• •	zůstanou stejné
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
1	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	43	0.06	22,12
24	44.4	38	4.6	0.02	
25	43.7	85	4.6	0.009	
26	41 4	95	4.1	0.005	
ź	58.8	54	4.8	0.027	
28	51.1	89	3.8	0.055	
9	61.6	65	4.1	0,000	
	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		
R	41,5	76	3.7		
	42.3	155	2,7		
ł	50.2	220			
╞	24.0	220			

Select variables to be transformed	& (optional) grouping variables	? 💌
1 - tlouštky 2 - studny 3 - letok tw 4 - kadr 5 - ovzdr V levém sloup	1 · tlouštky 2 · studny 3 · letokruhv ci vyberte	OK Stomo
položku "Str	ıdny"	Pro zobrazení odpovidají cích proměnných zvolte "Ukázat pouze". Pro
Vybrat vše Podrobn. Přiblížit Variables to transform:	Vybrat vše Podrobn. Přiblížit Grouping variable (optional):	stiskněte F1.
Pouze odpovídající proměnné		.41

- Houstky 23-letokruhy 4-kadmium 5-ovzduši	1 - tioustky 2 - studry 3 - letokruhy 5 - ovzduší m a pot stisknu	Highlight Bo wrd'te tím OK Pro zobrazeni odpovidajícich prománných zvote Ukázi poze Pro
Vybrat vše Podrobn. Přiblížit Variables to transform:	Vybrat vše Podrobn. Grouping variable (optional):	Přiblížit stiskněte F1.



🕂 Box-Cox Transformation: odhadyEDA1	
Box-Cox Image: Wariables: Box-Cox transformation Max. iterations: 40 Max. iterations: 40 Min. lambda -5 Max. lambda: 5 Epsilon (convergence): 00001 Shift variables with minimum<=0 to:	 ■ Po vložení ■ Po vložení ■ proměnné samotný výpočet spustíme ■ dalším klepnutím na tlačítko OK

Objeví se okno s výsledky transformace:
1) Summary - tabulka výsledků (hodnota λ 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 transformovana data
3) Search history plots (graf hledání řešení optimální hodnoty λ včetně vykreslení jejího IS)

🔊 Box-Cox Results: odhadyEDA1 ? 🗙 Box-Cox Results Summary SUHH Summary Cancel Histograms and normal probability plots 🤈 Options 🔻 1 Search history plots By Group Number of variables 1 Add variables to add: Write back to input spreadsheet.



Summary

Cancel

Options

Klepněte na "Search

Box-Cox Results

Number of variables to add:

Summary

Histograms and normal probability plots

Search history plots

11

SOHH







evropský sociální fond v ČR INVESTICE D	** MINISTERSTVO ŠKOLSTVÍ, KÁ UNIE MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY P DO ROZVOJE VZDĚLÁVÁNÍ	OP Vzdělávání ro konkurenceschopnost
Pomocí tlačítka v levém dolním rohu znovu vyvoláme menu výsledků Box-Cox Results odhadyEDAL Box-Cox Results odhadyEDAL Pripravena	Box-Cox Results: odhadyEDA1 Box-Cox Results Box-Cox Results Summarya klepněte na tlačítko "Summary" (Souhrn) s Write back to input spreadsheet	? × Summary Cancel ♪ Options ▼ ▶ Group

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"

STATISTICA Cz - PS 1* - [Box-Cox transformation (odhadyEDA1)] - [PS 1* - Box-Cox transformation (odhadyEDA1)]										
En Soubor Upravit Zobrazit Vloži	<u>F</u> ormát S <u>t</u> atistiky Data <u>m</u> ining <u>G</u> rafy <u>N</u> ástr	roje <u>D</u> ata <u>P</u> racovní sešit <u>O</u> kno Nápo <u>v</u> ěda								
0 🛩 🖬 🔁 🎒 🗛 👗 🖻 🕻	💅 🗠 🕬 🏘 Přidat do sešitu 👻 Přidat do	🛛 protokolu 👻 Přidat do MS Word 👻 🕼 🥔 🕽								
Arial 👻 10 👻	B I ∐ ≣ ≣ ⊒ 🗗 🗛 • 🏊 • 💿 •	' 🏷 🏢 👶 👯 🛤 😾 🐨 🖆 👌 🛃 **?								
🧀 PS 1*	Box-Cox transformati	ion (odbadyEDA1)								
🖃 🔤 (odhadyEDA1)	studny studny	ion (ounadyebAn)								
🗄 🔤 Box-Cox transformation	Transformed variable(s) original transform	ed								
Box-Cox transformation	1 32,0000 5,29	316								
Data stristics (odhadyEl	2 75.0000 7.37	533								
	3 16,0000 3,87	450								
	4 99,0000 8,14	673								
	5 80,0000 7,55	029								
	6 78,0000 7,48	135								
	7 28,0000 5,000	205								
	8 29,0000 5,07	769								
	9 170,0000 9,79	756								
	10 86,0000 7,74	946								
	11 80,0000 7,55	029								
	12 81,0000 7,58	427								
	13 26,0000 4,84	429								
	14 170,0000 9,79	756								
	15 81,0000 7,58	427								
	16 190,0000 10,16 17 005 0000 10,16	329								
	17 235,0000 10,88	8/1								
	10 30,0000 5,15	135								
	15 19,0000 4,20	550								
	16 190,0000 10,163 17 235,0000 10,883 18 30,0000 5,15 19 19,0000 4,203	329 871 135 550								

Hodnotu λ z listu Data Statistics 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 λ
- 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.)

X								
Sou	ibor Do	mů Vložen	í Rozlože	ení stránky	Vzorce	Data	Revize	Zobrazer
	📜 🔏 Vyjm	nout	Calibri	- 11	т А [±] . [•]	= _		-
	📃 📄 Кор	írovat 🔹	canon		A A	_		
VIo	žit 🍼 Kop	írovat formát	BIU	•	🏷 - <u>A</u> -	≣≣	•	Ne se s
	Schránk	a G		Písmo	Es.			Zarovná
	v) - (21 -	Ŧ						
	B8	• (f_s	r .				
	А	В	С	D	E		F	G
1								
				data	data			
2		hodnota 🤉	L	původní	transforme	ovaná		
3		0,229283	3	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				26,0,20	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		
_								



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 <u>o</u> blast: Sdružit: V Popisky v prvním řádk	\$D\$2:\$E\$101 Sloupce Řá<u>d</u>ky 		OK Storno Nápo <u>v</u> ěda
Možnosti výstupu Výs <u>t</u> upní oblast: Nový list: Nový s <u>e</u> šit	\$H\$3		
 Celkový přehled Hadina spolehlivosti pro stř K-té největší K-té nejmenší 	1 1] %]	

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:







N	data původní	data transformovaná
3		
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.







ŠKOLSTVÍ.

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 "svázány" 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 λ máme vypočítanou z programu STATISTICA):

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



Jediným problémem bude výpočet "lambda odmocniny". Známá 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

X												
Soubo	r Do	mů	Vlože	ní	Ro	zlože	ení s	trán	ky	V	zorc	e
ĥ	👗 Vyjm	out		Cali	bri			Ŧ	11	Ŧ	A	A
Vložit	🗳 Корі	írovat v	ormát	B	I	U	*		*	<u></u>	• 4	<u>+</u> -
	Schránk	a	- Fa				Pís	mo				- Fai
9	• (°' -	₹										
	J28		- (0		f,						
	А		В		С			D				Е

Dále nastavíme "Vybrat kategorii" na "Matematické" a objeví se následující okno

Argumenty fu	inkce
POWER	6
Číslo	= číslo
Exponent	E číslo
Umocní číslo na	= a zadaný exponent. Číslo je základ mocniny, kterým může být libovolné reálné číslo.
Výsledek =	
Nápověda k té	to funkci OK Storno

Do okna "Číslo" se vloží číslo pod odmocninou, do okna "Exponent" se vloží hodnota 1/λ.

Následující obrázek ukazuje vzorec pro výpočet retrasformovaného 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.







X												transforma	ce - Mic	rosoft Excel	
Sou	bor Dom	nů Vložení	Rozlože	ní stránky	Vzorce Data	Revi	ze Zo	brazení	í Doplňky						
	🖁 🔏 Vyjme	out	Calibri	• 18	• A A = =	=	≫~~	🚽 Za	alamovat text	Obecný	Ŧ		d	Normální_List1	l Normálni
Vlož	iit 🦪 Kopii	ovat ·	BIU		<u></u> → A - ≡ ≡	-	佳佳	•a• SI	oučit a zarovnat na střed 👻	· % 000	€,0 ,00	Podmíněné Forr	nátovat	Správně	Kontrolni
Ÿ	Schránka	ovaciormat		Bírmo			7.	vovnán	-	Číclo		formátování * jako i	abulku *		Chulu
	Schlanka	138		FISHIO	12		20	arovnan	1 (A)	CISIO	14				Styly
	POWER → (* X ✓ fx = POWER(JS*\$B\$3+1;1/\$B\$3)														
	Α	В	С	D	E	F		G	Н		1.1	J	K	L	M
				data	data										
2		hodnota λ		původní	transformovaná										
		0.229283		32.0000	5.29316						data	data			
3					-,						původní	transformovana			
4				75,0000	7,37533										
5				16,0000	3,87450				Stř. hodnota		93,05051	7,41919213	i		
6				99,0000	8,14673				Chyba stř. hodnoty		8,192695	0,199658333	5		
7				80,0000	7,55029				Medián		80	7,550292162	2		
8				78,0000	7,48135				Modus		80	7,550292162	2		
9				28,0000	5,00205				Směr. odchylka		81,51629	1,986575352	2		
10				29,0000	5,07769				Rozptyl výběru		6644,906	3,94648163	3		
11				170,0000	9,79756				Špičatost		20,29007	2,34591182	7		
12				86,0000	7,74946				Šikmost		3,869086	5 0,230096484	Ļ		
13				80,0000	7,55029				#ODKAZ!		597	7 13,2957008	3		
14				81,0000	7,58427				Minimum		3	3 1,249366603	5		
15				26,0000	4,84429				Maximum		600	14,5450674	1		
16				170,0000	9,79756				Součet		9212	2 734,5000213	3		
17				81,0000	7,58427				Počet		99	9 99)		
18				190,0000	10,16329				Hladina spolehlivosti (95,0%)	16,25814	0,396215468	3		
19				235,0000	10,88871										
20				30,0000	5,15135				dolní hranice IS		76,79237	7,02297666	/		
21				19,0000	4,20550				horní hranice IS		109,3086	5 7,81540760	1		
22				69,0000	7,15307										
23				79,0000	7,51599				střední hodnota retran	sformovaná		=POWER(.	5*\$E	3\$3+1;1/\$	B\$3)
24				105,0000	8,31663				dolní hranice IS retrans	formovaný					
25				63,0000	6,91539				horní hranice IS retrans	formovaný					
26				38,0000	5,68117										
27				95 0000	7 71702										

uata	data			
původní	transformovaná			
32,0000	5,29316		data původní	data transformovaná
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
	původní 32,0000 75,0000 16,0000 99,0000 80,0000 78,0000 28,0000 29,0000 170,0000 29,0000 170,0000 86,0000 81,0000 25,0000 170,0000 81,0000 235,0000 30,0000 19,0000 69,0000 79,0000 105,0000 63,0000	transformovaná původní transformovaná 32,0000 5,29316 75,0000 7,37533 16,0000 3,87450 99,0000 8,14673 80,0000 7,55029 78,0000 7,48135 28,0000 5,00205 29,0000 5,07769 170,0000 9,79756 86,0000 7,55029 81,0000 7,55029 81,0000 7,55029 81,0000 7,55029 81,0000 7,58427 26,0000 4,84429 170,0000 9,79756 81,0000 7,58427 190,0000 10,16329 235,0000 10,88871 30,0000 5,15135 19,0000 4,20550 69,0000 7,15307 79,0000 7,51599 105,0000 8,31663 63,0000 6,91539	uara uara původní transformovaná 32,000 5,29316 75,000 7,37533 16,000 3,87450 99,000 8,14673 Chyba stř. hodnota 99,000 8,14673 78,000 7,55029 Medián 78,000 7,48135 Modus 28,0000 5,00205 Směr. odchylka 29,0000 5,07769 Rozptyl výběru 170,000 9,79756 86,0000 7,74946 Sikmost 80,0000 7,55029 #ODKAZ! Minimum 26,0000 4,84429 Maximum 26,0000 170,000 9,79756 Součet 81,000 7,58427 Počet 190,000 10,16329 Hladina spolehlivosti (95,0%) 235,000 10,88871 30,0000 5,15135 dolní hranice IS 19,0000 4,20550	Value Value <th< td=""></th<>



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:

 $\underline{http://user.mendelu.cz/drapela/Statisticke_metody/Data_do_cviceni/Statistica/odhadyEDA1.st} \underline{a}$

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" odlehlých hodnot z důvodů grafické přehlednosti možné.

Stanovení hodnoty λ:





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).







MLÁDEŽE A TĚLOV



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
Rozpty	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 λ je 0,295 s intervalem spolehlivosti (0.016; 0.612). Z toho vyplývá, že transformace může být úspěšná (IS λ 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	26,86	4,84
dolní hranice IS stř. hodnoty	15,53	3,78
horní hranice IS stř. hodnoty	38,18	5,90
střední hodnota		20.17
retransformovaná		20,17
dolní hranice IS stř. hodnoty	N	13.64
retransformovaná	2	12,04
horní hranice IS stř. hodnoty		20.44
retransformovaná		50,41

Grafické porovnání IS pro původní a transformovaná data je v prezentaci

http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/BoxCox_transformace ovzdusi.ppsx

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:







 $\frac{https://support.microsoft.com/cs-cz/office/zobrazen\%C3\%AD-prezentace-bez-powerpointu-2f1077ab-9a4e-41ba-9f75-d55bd9b231a6}{2f1077ab-9a4e-41ba-9f75-d55bd9b231a6}$