

ÚPLNÝ FAKTORIÁLNY DIZAJN S MERANÍM V CENTRÁLNOBODE – RIEŠENÝ VZOROVÝ PRÍKLAD

Bola skúmaná závislosť rýchlosti nitridového leptania plazmových dosťičiek v závislosti od vybraných premenných: medzery medzi elektródami, prietoku plynu a výkonu na katóde, pričom $x_1 \in \langle 0,8; 1,2 \rangle$ je medzera medzi elektródami v cm, $x_2 \in \langle 125; 200 \rangle$ je prietok plynu v cm^3/min , $x_3 \in \langle 275; 325 \rangle$ je výkon na katóde vo W. Zadanie príkladu a namerané hodnoty boli prevzaté z Montgomery, str. 245.

Pri uvažovanom experimente typu 2^3 boli namerané hodnoty v 10^{-10} m/min:

$f(0,8;125;325)=1052$; $f(0,8;125;275)=550$; $f(0,8;200;275)=601$; $f(1,2;200;275)=635$;
 $f(1,2;125;275)=669$; $f(0,8;200;275)=633$; $f(1,2;125;275)=650$; $f(0,8;125;275)=604$;
 $f(1,2;200;275)=642$; $f(0,8;125;325)=1037$; $f(1,2;125;325)=749$; $f(1,2;200;325)=860$;
 $f(1,2;125;325)=868$; $f(0,8;200;325)=1075$; $f(1,2;200;325)=729$; $f(0,8;200;325)=1063$

Pri dvojúrovňovom experimente bola závislá premenná (odozva) – rýchlosť nitridového leptania pre vstupné nezávislé premenné (faktory) vždy meraná na dolnej úrovni, t.j. pri nastavenej minimálnej hodnote a hornej úrovni, t.j. pri maximálnej hodnote ako uvádza tabuľka 1.

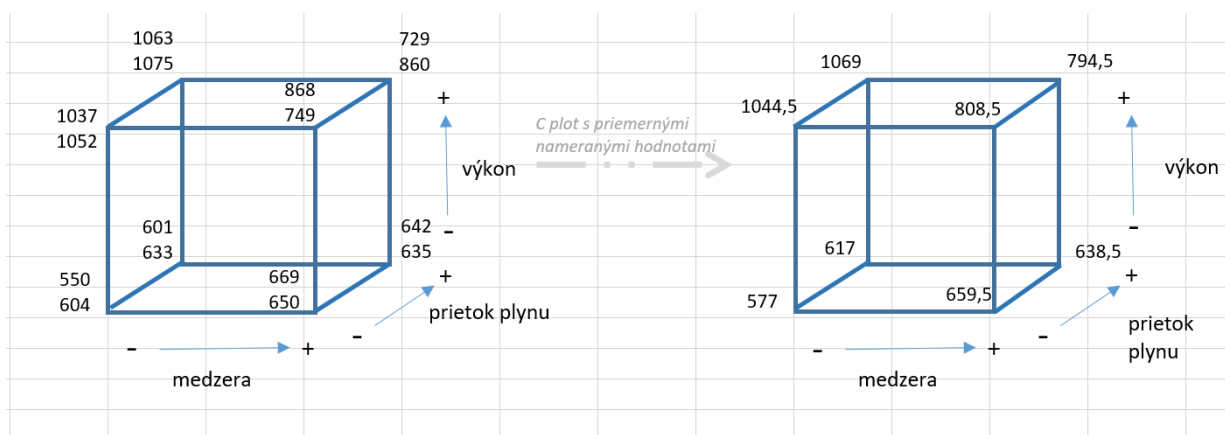
Tabuľka 1 Vstupné nezávislé faktory a ich úrovne				
Faktory	v pôvodných jednotkách		v kódovaných jednotkách	
	dolná úroveň	horná úroveň	dolná úroveň	horná úroveň
A_medzera medzi elektródami	0,8 cm	1,2 cm	-1	1
B_prietok plynu	125 cm^3/min	200 cm^3/min	-1	1
C_výkon na katóde	275 W	325 W	-1	1

Zápis nameraných hodnôt a lineárna transformácia

Počet bodov experimentálneho priestoru, v ktorých bolo realizované meranie je pri úplnom faktoriálnom experimente typu 2^3 rovný 8. Vzhľadom na 16 nameraných hodnôt a predpokladajúc, že experiment bol vyvážený, boli pri experimente použité 2 replikácie. Prehľadný zápis nameraných hodnôt zodpovedajúci rôznym kombináciám úrovní faktov je v tabuľke 2.

Tabuľka 2 Nitridové leptanie, namerané hodnoty						
	t_1	t_2	t_3	$y_1 [10^{-10} \text{ m/min}]$	$y_2 [10^{-10} \text{ m/min}]$	$\bar{y} [10^{-10} \text{ m/min}]$
(1)	-1	-1	-1	550	604	577
a	1	-1	-1	669	650	659,5
b	-1	1	-1	601	663	617
c	-1	-1	1	1037	1052	1044,5
ab	1	1	-1	642	635	638,5
ac	1	-1	1	868	749	808,5
bc	-1	1	1	1063	1075	1069
abc	1	1	1	729	860	794,5

Cplot na obr. 1.1 s nameranými hodnotami a ich vypočítanou priemernou hodnotou v jednotlivých experimentálnych bodoch poskytuje vizuálnu predstavu o výsledkoch experimentu vzhľadom na nastavené úrovne faktorov.



Obr. 1.1 Grafické znázornenie nameraných hodnôt a ich priemerov v jednotlivých experimentálnych bodoch pri experimente typu 2^3

Stredy intervalov s a polovičné dĺžky variačného rozpätia intervalov λ na nasledujúcom obr. 1.2 sú použitý pri kódovaní všetkých úrovní vstupných hodnôt na obr.1.3.

	T	U	V
1		s	lambda
2	x_1	$=(0,8+1,2)/2$	$=(1,2-0,8)/2$
3	x_2	$=(125+200)/2$	$=(200-125)/2$
4	x_3	$=(275+325)/2$	$=(325-275)/2$

	T	U	V
1		s	lambda
2	x_1	1	0,2
3	x_2	162,5	37,5
4	x_3	300	25

Obr. 1.2 Výpočet stredy a polovice variačného rozpätia intervalov faktorov x_1 , x_2 a x_3

	J	K	L	M	N	O	P	Q	R	S
1	Transformácia:									
2		x1	x2	x3	Y		t1	t2	t3	Y
3	1	0,8	125	325	1052		=(K3-\$U\$2)/\$V\$2	=(L3-\$U\$3)/\$V\$3	=(M3-\$U\$4)/\$V\$4	1052
4	2	0,8	125	275	550		=(K4-\$U\$2)/\$V\$2	=(L4-\$U\$3)/\$V\$3	=(M4-\$U\$4)/\$V\$4	550
5	3	0,8	200	275	601		=(K5-\$U\$2)/\$V\$2	=(L5-\$U\$3)/\$V\$3	=(M5-\$U\$4)/\$V\$4	601
6	4	1,2	200	275	635		=(K6-\$U\$2)/\$V\$2	=(L6-\$U\$3)/\$V\$3	=(M6-\$U\$4)/\$V\$4	635
7	5	1,2	125	275	669		=(K7-\$U\$2)/\$V\$2	=(L7-\$U\$3)/\$V\$3	=(M7-\$U\$4)/\$V\$4	669
8	6	0,8	200	275	633		=(K8-\$U\$2)/\$V\$2	=(L8-\$U\$3)/\$V\$3	=(M8-\$U\$4)/\$V\$4	633
9	7	1,2	125	275	650		=(K9-\$U\$2)/\$V\$2	=(L9-\$U\$3)/\$V\$3	=(M9-\$U\$4)/\$V\$4	650
10	8	0,8	125	275	604		=(K10-\$U\$2)/\$V\$2	=(L10-\$U\$3)/\$V\$3	=(M10-\$U\$4)/\$V\$4	604
11	9	1,2	200	275	642		=(K11-\$U\$2)/\$V\$2	=(L11-\$U\$3)/\$V\$3	=(M11-\$U\$4)/\$V\$4	642
12	10	0,8	125	325	1037		=(K12-\$U\$2)/\$V\$2	=(L12-\$U\$3)/\$V\$3	=(M12-\$U\$4)/\$V\$4	1037
13	11	1,2	125	325	749		=(K13-\$U\$2)/\$V\$2	=(L13-\$U\$3)/\$V\$3	=(M13-\$U\$4)/\$V\$4	749
14	12	1,2	200	325	860		=(K14-\$U\$2)/\$V\$2	=(L14-\$U\$3)/\$V\$3	=(M14-\$U\$4)/\$V\$4	860
15	13	1,2	125	325	868		=(K15-\$U\$2)/\$V\$2	=(L15-\$U\$3)/\$V\$3	=(M15-\$U\$4)/\$V\$4	868
16	14	0,8	200	325	1075		=(K16-\$U\$2)/\$V\$2	=(L16-\$U\$3)/\$V\$3	=(M16-\$U\$4)/\$V\$4	1075
17	15	0,8	200	325	1063		=(K17-\$U\$2)/\$V\$2	=(L17-\$U\$3)/\$V\$3	=(M17-\$U\$4)/\$V\$4	1063
18	16	1,2	200	325	729		=(K18-\$U\$2)/\$V\$2	=(L18-\$U\$3)/\$V\$3	=(M18-\$U\$4)/\$V\$4	729

	J	K	L	M	N	O	P	Q	R	S
1	Transformácia:									
2		x1	x2	x3	Y		t1	t2	t3	Y
3	1	0,8	125	325	1052		-1	-1	1	1052
4	2	0,8	125	275	550		-1	-1	-1	550
5	3	0,8	200	275	601		-1	1	-1	601
6	4	1,2	200	275	635		1	1	-1	635
7	5	1,2	125	275	669		1	-1	-1	669
8	6	0,8	200	275	633		-1	1	-1	633
9	7	1,2	125	275	650		1	-1	-1	650
10	8	0,8	125	275	604		-1	-1	-1	604
11	9	1,2	200	275	642		1	1	-1	642
12	10	0,8	125	325	1037		-1	-1	1	1037
13	11	1,2	125	325	749		1	-1	1	749
14	12	1,2	200	325	860		1	1	1	860
15	13	1,2	125	325	868		1	-1	1	868
16	14	0,8	200	325	1075		-1	1	1	1075
17	15	0,8	200	325	1063		-1	1	1	1063
18	16	1,2	200	325	729		1	1	1	729

Obr. 1.3 Transformácia pôvodných úrovní faktorov do kódovaných hodnôt realizovaná v Exceli

Výpočet bodových odhadov koeficientov regresného modelu MNŠ

Koeficienty $\hat{B}$ pre regresný model lineárneho typu $y = B_0 + \sum_{j=1}^k B_j x_j + \sum \sum_{i < j} B_{ij} x_i x_j$ odhadujeme pomocou $\hat{b}$ -koeficientov transformovaného regresného modelu, t.j. pomocou transformovanej regresnej funkcie lineárneho typu $(\hat{f}_L)_t(t_1, t_2, t_3) = b_0 + \sum_{j=1}^k b_j t_j + \sum \sum_{i < j} b_{ij} t_i t_j$. Pri využití MNŠ použijeme vzťah

$$\hat{b} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_{12} \\ \hat{b}_{13} \\ \hat{b}_{23} \\ \hat{b}_{123} \end{pmatrix} = \frac{1}{m \cdot 2^k} X^T Y$$

Pri výpočte v Exceli vytvoríme maticu plánu experimentu X z pôvodných kódovaných hodnôt na obr. 1.4 pridaním stĺpca 1 pri b_0 a stĺpcov, ktoré reprezentujú interakcie faktorov v modeli. Súčin transponovanej matice plánu experimentu X^T a matice nameraných hodnôt zodpovedajúcich príslušným úrovniam faktorov Y delíme počtom nameraných hodnôt $m \cdot 2^k$ na obr. 1.5.

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
1			t1	t2	t3	t1*t2	t1*t3	t2*t3	t1*t2*t3								
2		b0	b1	b2	b3	b12	b13	b23	b123		Y						
3	Matica plánu experimentu X	1	-1	-1	1	1	-1	-1	1		1052						
4		1	-1	-1	-1	1	1	1	-1		550						
5		1	-1	1	-1	-1	1	-1	1		601						
6		1	1	1	-1	1	-1	-1	-1		635						
7		1	1	-1	-1	-1	-1	1	1		669						
8		1	-1	1	-1	-1	1	-1	1		633						
9		1	1	-1	-1	-1	-1	1	1		650						
10		1	-1	-1	-1	1	1	1	-1		604						
11		1	1	1	-1	1	-1	-1	-1		642						
12		1	-1	-1	1	1	-1	-1	1		1037						
13		1	1	-1	1	-1	1	-1	-1		749						
14		1	1	1	1	1	1	1	1		860						
15		1	1	-1	1	-1	1	-1	-1		868						
16		1	-1	1	1	-1	-1	1	-1		1075						
17		1	-1	1	1	-1	-1	1	-1		1063						
18		1	1	1	1	1	1	1	1		729						
19																	
20	XT=	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21		-1	-1	-1	1	1	-1	1	-1	1	-1	1	1	1	-1	-1	1
22		-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	1
23		1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1
24		1	1	-1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1
25		-1	1	1	-1	-1	1	-1	1	-1	-1	1	1	1	-1	-1	1
26		-1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	-1	1	1	1
27		1	-1	1	-1	1	1	1	-1	-1	1	-1	1	-1	-1	-1	1
28																	
29																	
30	XT*Y=	12417		b ₀ =	776,0625												
31		-813		b ₁ =	-50,8125												
32		59		b ₂ =	3,6875												
33		2449		b ₃ =	153,0625												
34		-199		b ₁₂ =	-12,4375												
35		-1229		b ₁₃ =	-76,8125												
36		-17		b ₂₃ =	-1,0625												
37		45		b ₁₂₃ =	2,8125												

Obr. 1.4 Matica plánu experimentu X potrebná na výpočet bodových odhadov koeficientov regresného modelu

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	
1			t1	t2	t3	t1*t2	t1*t3	t2*t3	t1*t2*t3		
2		b0	b1	b2	b3	b12	b13	b23	b123		Y
3	X=	1	-1	-1	1	=AA3*AB3	=AA3*AC3	=AB3*AC3	=AA3*AB3*AC3		1052
4		1	-1	-1	-1	=AA4*AB4	=AA4*AC4	=AB4*AC4	=AA4*AB4*AC4		550
5		1	-1	1	-1	=AA5*AB5	=AA5*AC5	=AB5*AC5	=AA5*AB5*AC5		601
6		1	1	1	-1	=AA6*AB6	=AA6*AC6	=AB6*AC6	=AA6*AB6*AC6		635
7		1	1	-1	-1	=AA7*AB7	=AA7*AC7	=AB7*AC7	=AA7*AB7*AC7		669
8		1	-1	1	-1	=AA8*AB8	=AA8*AC8	=AB8*AC8	=AA8*AB8*AC8		633
9		1	1	-1	-1	=AA9*AB9	=AA9*AC9	=AB9*AC9	=AA9*AB9*AC9		650
10		1	-1	-1	-1	=AA10*AB10	=AA10*AC10	=AB10*AC10	=AA10*AB10*AC10		604
11		1	1	1	-1	=AA11*AB11	=AA11*AC11	=AB11*AC11	=AA11*AB11*AC11		642
12		1	-1	-1	1	=AA12*AB12	=AA12*AC12	=AB12*AC12	=AA12*AB12*AC12		1037
13		1	1	-1	1	=AA13*AB13	=AA13*AC13	=AB13*AC13	=AA13*AB13*AC13		749
14		1	1	1	1	=AA14*AB14	=AA14*AC14	=AB14*AC14	=AA14*AB14*AC14		860
15		1	1	-1	1	=AA15*AB15	=AA15*AC15	=AB15*AC15	=AA15*AB15*AC15		868
16		1	-1	1	1	=AA16*AB16	=AA16*AC16	=AB16*AC16	=AA16*AB16*AC16		1075
17		1	-1	1	1	=AA17*AB17	=AA17*AC17	=AB17*AC17	=AA17*AB17*AC17		1063
18		1	1	1	1	=AA18*AB18	=AA18*AC18	=AB18*AC18	=AA18*AB18*AC18		729
19											
20	XT=	=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
21		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
22		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
23		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
24		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
25		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
26		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
27		=TRANSPOSE(Z3:AG18)	=TRANS	=TRANS	=TRANS	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANSPOSE(Z3:AG1	=TRANS
28											
29											
30	XT*Y=	=MMULT(Z20:AO27;A13:A118)				b ₀ =	=Z30/16				
31		=MMULT(Z20:AO27;A13:A118)				b ₁ =	=Z31/16				
32		=MMULT(Z20:AO27;A13:A118)				b ₂ =	=Z32/16				
33		=MMULT(Z20:AO27;A13:A118)				b ₃ =	=Z33/16				
34		=MMULT(Z20:AO27;A13:A118)				b ₁₂ =	=Z34/16				
35		=MMULT(Z20:AO27;A13:A118)				b ₁₃ =	=Z35/16				
36		=MMULT(Z20:AO27;A13:A118)				b ₂₃ =	=Z36/16				
37		=MMULT(Z20:AO27;A13:A118)				b ₁₂₃ =	=Z37/16				

Po označení výslednej veľkosti poľa
potvrdzujeme trojkombináciou kláves
CTRL+ Shift + Enter

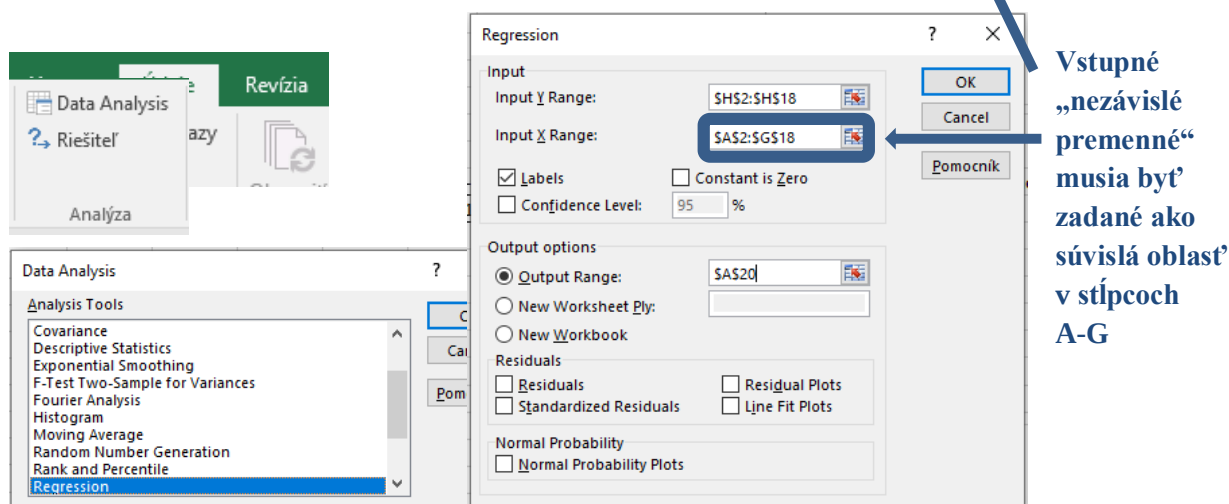
Obr. 1.5 Výpočet bodových odhadov koeficientov regresného modelu realizovaný v Exceli

Hierarchický regresný model a identifikácia signifikantných faktorov a interakcií faktorov

V prvej fáze vyhodnocovania experimentu nás zvyčajne zaujíma vplyv ktorých faktorov, resp. interakcií možno považovať za signifikantný a ktorý možno zanedbať. Bodové odhady regresných koeficientov, hodnoty testovacích štatistík, p-hodnoty a intervaly spoľahlivosti, vieme jednoducho získať cez analýzu dát v Exceli postupnosťou príkazov *Data→Data analysis→Regression*.

Vstupné údaje, ktoré použijeme v regresnej analýze sú na obr. 1.6, pričom ako nezávislá premenná X musí byť zapísaná súvislá oblasť obsahujúca stĺpce predstavujúce všetky členy, ktoré majú v regresnom modeli vystupovať, t.j. všetky faktory a ich interakcie vyjadrené v kódovaných jednotkách. Ako závislú premennú Y zadávame namerané hodnoty odozvy v príslušnom poradí kombinácií úrovní vstupov. Ak označíme aj riadok, v ktorom sa nachádzajú označenia príslušného faktora alebo interakcie, potom je potrebné v dialógovom okne potvrdiť *Labels*.

	A	B	C	D	E	F	G	H
1								
2	t1	t2	t3	t1*t2	t1*t3	t2*t3	t1*t2*t3	yi
3	-1	-1	-1	1	1	1	-1	550
4	-1	-1	-1	1	1	1	-1	604
5	1	-1	-1	-1	-1	1	1	669
6	1	-1	-1	-1	-1	1	1	650
7	-1	1	-1	-1	1	-1	1	633
8	-1	1	-1	-1	1	-1	1	601
9	1	1	-1	1	-1	-1	-1	642
10	1	1	-1	1	-1	-1	-1	635
11	-1	-1	1	1	-1	-1	1	1037
12	-1	-1	1	1	-1	-1	1	1052
13	1	-1	1	-1	1	-1	-1	749
14	1	-1	1	-1	1	-1	-1	868
15	-1	1	1	-1	-1	1	-1	1075
16	-1	1	1	-1	-1	1	-1	1063
17	1	1	1	1	1	1	1	729
18	1	1	1	1	1	1	1	860



Obr. 1.6 Zadávanie vstupných údajov použité pri regresnej analýze s viacerými vstupnými parametrami v Exceli

Výstup regresnej analýzy v Exceli je na obr. 1.7 a 1.8.

Regression Statistics					
Multiple R	0,982898759				
R Square	0,96608997				
Adjusted R Square	0,936418694				
Standard Error	47,46116834				
Observations	16				

Obr. 1.7 Výstup regresnej analýzy v Exceli – všeobecné charakteristiky týkajúce sa nájdeného regresného modelu

ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	7	513400,4375	73342,9	32,5598	2,89617E-05	
Residual	8	18020,5	2252,56			
Total	15	531420,9375				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	776,0625	11,86529208	65,4061	3,3E-12	748,7010874	803,42391
t1	-50,8125	11,86529208	-4,28245	0,00268	-78,1739126	-23,45109
t2	3,6875	11,86529208	0,31078	0,76391	-23,6739126	31,048913
t3	153,0625	11,86529208	12,9	1,2E-06	125,7010874	180,42391
t1*t2	-12,4375	11,86529208	-1,04823	0,32517	-39,7989126	14,923913
t1*t3	-76,8125	11,86529208	-6,47371	0,00019	-104,173913	-49,45109
t2*t3	-1,0625	11,86529208	-0,08955	0,93085	-28,4239126	26,298913
t1*t2*t3	2,8125	11,86529208	0,23704	0,81859	-24,5489126	30,173913

Obr. 1.8 Výstup regresnej analýzy v Exceli – tabuľka ANOVA a tabuľka so štatistikami týkajúcimi sa regresných koeficientov

Stĺpec *Coefficients* obsahuje bodové odhady $\hat{b}$ koeficientov transformovanej regresnej funkcie lineárneho typu $f_t(t_1, t_2, t_3)$: $\hat{b}_0 = 776,0625$ (číselná hodnota zodpovedá aritmetickému priemeru všetkých nameraných hodnôt Y), $\hat{b}_1 = -50,8125$, $\hat{b}_2 = 3,6875$, $\hat{b}_3 = 153,0625$, $\hat{b}_{12} = -12,4375$, $\hat{b}_{13} = -76,8125$, $\hat{b}_{23} = -1,0625$, $\hat{b}_{123} = 2,8125$. Sú to koeficienty, ktoré sa vyskytujú pri faktoroch a interakciách faktorov v regresnom modeli $(f_L)_t(t_1, t_2, t_3) = b_0 + b_1t_1 + b_2t_2 + b_3t_3 + b_{12}t_1t_2 + b_{13}t_1t_3 + b_{23}t_2t_3 + b_{123}t_1t_2t_3$.

Pri rozhodovaní o ďalšom ponechaní, resp. vylúčení faktorov v ďalšom skúmaní môžeme napríklad pomocou p-hodnoty v stĺpci *P-value* identifikovať, ktoré koeficienty, a tým aj ktoré členy modelu v ňom majú zostať a ktoré môžeme vylúčiť. Rozhodneme vplyv ktorých faktorov či interakcií je, či nie je štatisticky významný (signifikantný) vzhľadom na skúmanú rýchlosť nitridového leptania. V našom prípade možno na hladine významnosti 0,05 možno vylúčiť faktor prietok plynu spolu so všetkými interakciami s týmto faktorom, pretože p-hodnota je väčšia ako nami zvolené $\alpha = 0,05$. K rovnakému výsledku by sme sa dopracovali aj pri rozhodovaní na základe intervalov spoľahlivosti. Pri každom z vylúčených členov modelu nájdeme interval spoľahlivosti obsahujúci 0. Po vylúčení nevýznamných členov z pôvodného regresného modelu dostaneme regresný model, ktorý spĺňa podmienku hierarchickosti - ak v modeli zostáva člen, ktorý obsahuje interakcie, ponechávame v modeli aj všetky členy, ktoré v interakcii vystupujú bez ohľadu na ich štatistickú významnosť.

V prípade riešenia pomocou softvéru Minitab volíme metódu postupnej eliminácie (Backward elimination), kedy postupne z modelu vylúčime členy štatisticky nevýznamné na zvolenej hladine významnosti.

Po opätovnom spustení regresnej analýzy dostávame nové aktuálne hodnoty R^2 , ANOVY a testovacie štatistiky týkajúce sa členov regresného modelu. Hodnoty regresných koeficientov ponechaných členov zostávajú rovnaké nakoľko s výnimkou úplného kvadratického modelu sú vždy rovné polovici veľkosti efektov.

Pri ďalšom vyhodnotení z dôvodu prepočítania výsledkov je potrebné regresnú analýzu spustiť opakovane. Ponechané vstupy musia byť v Exceli uvedené ako súvislá oblasť, viď obr. 1.9. Výstup regresnej analýzy je na obr. 1.10.

t1	t3	t1*t3	yi
-1	-1	1	550
-1	-1	1	604
1	-1	-1	669
1	-1	-1	650
-1	-1	1	633
-1	-1	1	601
1	-1	-1	642
1	-1	-1	635
-1	1	-1	1037
-1	1	-1	1052
1	1	1	749
1	1	1	868
-1	1	-1	1075
-1	1	-1	1063
1	1	1	729
1	1	1	860

Obr. 1.9 Pri ďalšom vyhodnotení z dôvodu správnych výsledkov je potrebné regresnú analýzu spustiť opakovane. Ponechané vstupy musia byť v Exceli opäť uvedené ako súvislá oblasť.

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0,980179					
R Square	0,960751					
Adjusted R Squa	0,950939					
Standard Error	41,69108					
Observations	16					
ANOVA						
	df	SS	MS	F	gnificance F	
Regression	3	510563,2	170187,7	97,91338	1,05E-08	
Residual	12	20857,75	1738,146			
Total	15	531420,9				
	Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	776,0625	10,42277	74,45838	2,29E-17	753,3532	798,7718
t1	-50,8125	10,42277	-4,87514	0,000382	-73,5218	-28,1032
t3	153,0625	10,42277	14,6854	4,95E-09	130,3532	175,7718
t1*t3	-76,8125	10,42277	-7,36968	8,62E-06	-99,5218	-54,1032

Obr. 1.10 Výstup regresnej analýzy v Exceli po vylúčení nevýznamných faktorov a interakcií

Bodový odhad transformovanej regresnej funkcie lineárneho typu $(\hat{f}_L)_t(t_1, t_2, t_3)$, ktorú dostaneme po vylúčení štatisticky nevýznamných členov je

$$(\hat{f}_L)_t(t_1, t_3) = 776,0625 - 50,8125t_1 + 153,0625t_3 - 76,8125t_1t_3$$

Odtiaľ vieme vyjadriť bodový odhad regresnej funkcie lineárneho typu $\hat{f}_L(x_1, x_2, x_3)$, ktorá bude v pôvodných jednotkách. V neupravenom tvare po spätnom dosadení transformačných vzťahov

$$\begin{aligned} \hat{f}_L(x_1, x_3) \\ = 776,0625 - 50,8125 \frac{x_1 - 1}{0,2} + 153,0625 \frac{x_3 - 300}{25} - 76,8125 \frac{x_1 - 1}{0,2} \cdot \frac{x_3 - 300}{25} \end{aligned}$$

Po úprave

$$\hat{f}_L(x_1, x_3) = -5415,375 + 4354,688x_1 + 21,485x_3 - 15,3625x_1x_3$$

Individuálne t-testy štatistickej významnosti regresných koeficientov

Má medzera medzi elektródami výrazný vplyv na rýchlosť nitridového leptania? Použijeme individuálne t-testy štatistickej významnosti regresných koeficientov. V princípe si môžeme zvoliť ľubovoľnú malú chybu I.druhu α , ktorej sa môžeme dopustiť pri vyhodnocovaní hypotéz. Najčastejšie sa v praxi volí $\alpha = 0,05$, resp. 0,01 alebo 0,1. Pre naše účely bude z dôvodu názornosti výpočtov volená vždy iná malá hodnota α . Napr. na hladine významnosti 3% chceme otestovať, či je samostatné pôsobenie faktora x_1 – medzera medzi elektródami, na hodnotu regresnej funkcie lineárneho typu štatisticky významné.

Pri testovaní vždy používame kódované hodnoty, resp. transformovanú regresnú funkciu, preto samostatné pôsobenie faktora x_1 na hodnotu $f_L(x_1, x_3)$ vyhodnocujeme ako samostatné pôsobenie transformovaného faktora t_1 na hodnotu $(f_L)_t(t_1, t_3) = b_0 + b_1t_1 + b_3t_3 + b_{13}t_1t_3$. Samostatné pôsobenie transformovaného faktora t_1 testujeme pomocou koeficientu b_1 . Nulová hypotéza H_0 hovorí, že regresný koeficient b_1 nie je štatisticky významný, oproti stojí obojstranná alternatíva H_1 regresný koeficient b_1 je štatisticky významný: $H_0: b_1 = 0 \leq H_1: b_1 \neq 0$.

Z výstupu Excelu na obr.1.11 použijeme riadok, v ktorom sa nachádzajú údaje súvisiace s koeficientom b_1 .

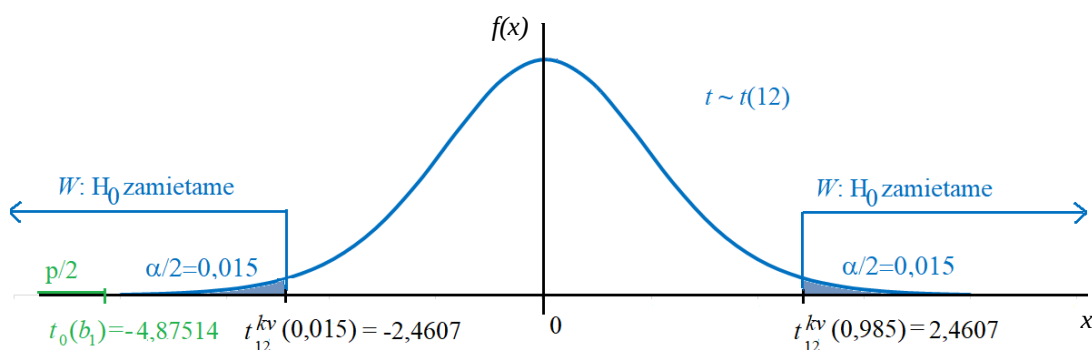
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 97,0%	Upper 97,0%
Intercept	776,0625	10,42276905	74,45838	2,29E-17	753,353237	798,771763	750,4151905	801,7098095
t1	-50,8125	10,42276905	-4,87514	0,000382	-73,521763	-28,1032371	-76,4598095	-25,16519048
t3	153,0625	10,42276905	14,6854	4,95E-09	130,353237	175,771763	127,4151905	178,7098095
t1*t3	-76,8125	10,42276905	-7,36968	8,62E-06	-99,521763	-54,1032371	-102,45981	-51,16519048

Obr. 1.11 Tabuľka s bodovými odhadmi koeficientov a im prislúchajúcimi hodnotami

Testovacia charakteristika (štatistika) $t(b_1) = \frac{\hat{b}_1}{SE(b_1)} \sim t(n - k - 1)$, kde $\hat{b}_1$ je bodový odhad regresného koeficientu b_1 : $\hat{b}_1 = -50,8125$, $SE(b_1) = 10,42276905$ je štandardná chyba tohto odhadu, $n = 16$ je celkový počet nameraných hodnôt, $k = 3$ je počet členov testovaného regresného modelu (členy b_1t_1 , b_3t_3 a $b_{13}t_1t_3$). Odtiaľ $n - k - 1 = 16 - 3 - 1 = 12$. Parameter Studentovho rozdelenia vyjadrený podľa vzťahu $(m - 1) \cdot 2^k$ je rovný $(4 - 1) \cdot 2^2 = 12$, nakoľko vylúčením faktora prietok plynu sa experiment zredukoval na experiment typu 2^2 , pričom počet replikácií pri zachovaní celkového počtu nameraných hodnôt sa zvýšil na 4. Zápis $t(b_i) \sim t(12)$ čítame: testovacie štatistiky pre koeficienty b_i majú Studentovo rozdelenie s 12-timi stupňami voľnosti.

Na základe analyzovaných údajov vypočítaná hodnota testovacej štatistiky, t.j. realizácia testovacej štatistiky $t_0(b_1) = \frac{-50,8125}{10,42276905} \doteq -4,87514$. Túto hodnotu nájdeme v stĺpci *t Stat*.

Vizuálnu predstavu o rozdelení testovacej štatistiky $t(b_1)$, o umiestnení hodnoty testovacej štatistiky t_0 vzhľadom na kritickú oblasť W zamietnutia nulovej hypotézy H_0 , ktorú sme určili podľa zvolenej hladiny významnosti α , o veľkosti p-hodnoty v porovnaní s hodnotou α dáva nasledujúci obr. 1.12



Obr. 1.12 Grafické riešenie pre testovanie hypotézy o koeficiente b_1

O výsledku testu hypotéz rozhodujeme 3 spôsobmi, pričom vždy musíme dostať rovnaký záver:

- Pomocou kritickej oblasti (kritického oboru) $W = (-\infty; -2,4607) \cup (2,4607; \infty)$, kde kritickú hodnotu určíme ako 98,5% kvantil $t_{12}^{kv}(0,985) = T.INV(0,985; 12) = 2,4607$.
1. možnosť vyhodnotenia: $|t_0(b_1)| = 4,87514 > t_{12}^{kv}(0,985)$ alebo
2. možnosť vyhodnotenia: realizácia testovacej štatistiky $t_0(b_1) = -4,87514 \in W$, preto zamietame $H_0 : b_1 = 0$, viď obr. 1.12.

b) Vyhodnotenie pomocou p-hodnoty

Túto hodnotu nájdeme v stĺpci *P-value*. Keďže p-hodnota predstavuje pravdepodobnosť, že dostaneme rovnaký alebo ešte extrémnejší výsledok v prípade, že by platila nulová hypotéza H_0 , jej číselnú hodnotu vieme vypočítať pomocou distribučnej funkcie.

$$p(b_1) = 2 * \min\{F(t_0(b_1)); 1 - F(t_0(b_1))\} = 0,000382$$

Výpočet p-hodnoty v Exceli je na obr. 1.13.

F(t0)=	=T.DIST(-4,87514;12;1)	F(t0)=	0,00019082
1-F(t0)=	=1-S63	1-F(t0)=	0,99980918
p=	=2*MIN(S63:S64)	p=	0,00038164

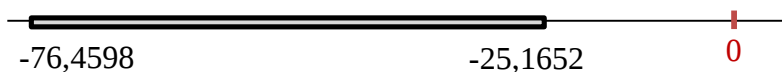
Obr. 1.13 Výpočet p-hodnoty pomocou distribučnej funkcie

V prípade p-hodnoty nepotrebujeme kritickú oblasť W , t.j. nie je potrebný ani výpočet kvantilu $t_{12}^{kv} \left(1 - \frac{\alpha}{2}\right)$, pretože p-hodnotu porovnávame so zvolenou hladinou významnosti α . Čím je p-hodnota menšia, tým viac sme presvedčení, že nulovú hypotézu treba zamietnuť.

$$p = 0,000382 < 0,03 = \alpha, \text{ preto zamietame } H_0 : b_1 = 0$$

c) Vyhodnotenie pomocou $(1 - \alpha) \cdot 100\%$ intervalu spoľahlivosti, alebo tzv. konfidenčného intervalu.

Pri vyhodnotení testov hypotéz pomocou intervalu spoľahlivosti o prijatí alebo zamietnutí nulovej hypotézy rozhodujeme na základe toho, či 0 patrí alebo nepatrí do $(1 - \alpha) \cdot 100\%$ IS. V našom prípade $\alpha = 0,03$, preto potrebujeme 97% IS pre koeficient b_1 . Dolnú hranicu intervalu nájdeme vo výstupe regresnej analýzy v tabuľke na obr. 1.11 v stĺpci *Lower 97%*, hornú hranicu v stĺpci *Upper 97%*. 97% IS pre koeficient b_1 : (-76,4598; -25,1652)



Obr. 1.14 0 nepatrí do 97% IS pre koeficient b_1

Na výpočet hraníc použijeme vzťah $b_1 \pm \Delta$, kde $\Delta = SE(b_1) \cdot t_{(m-1), 2k}^{kv} \left(1 - \frac{\alpha}{2}\right)$, obr. 1.15.

	I	J		I	J
55			55		
56	b1=	-50,8125	56	b1=	-50,8125
57	SE(b1)=	10,4227690458598	57	SE(b1)=	10,42277
58	t kv 12 (0,985)=	=T.INV(1-0,03/2;12)	58	t kv 12 (0,985)=	2,4607
59	delta=	=J57*J58	59	delta=	25,64731
60	b1-delta=	=J56-J59	60	b1-delta=	-76,4598
61	b1+delta=	=J56+J59	61	b1+delta=	-25,1652

Obr. 1.15 Výpočet intervalu spoľahlivosti pre koeficient b_1 v Exceli

Nulová hypotéza $H_0 : b_1 = 0$ hovorí, že regresný koeficient b_1 je nulový, preto sme overovali či 0 patrí alebo nepatrí do daného konfidenčného intervalu.

Keďže $0 \notin (-76,4598; -25,1652)$, so spoľahlivosťou 97% môžeme hypotézu $H_0 : b_1 = 0$ zamietnuť.

Záver:

Na hladine významnosti $\alpha = 0,03$ možno konštatovať, že regresný koeficient b_1 je štatisticky významný, t.j. samostatný vplyv faktora t_1 na hodnotu transformovanej regresnej funkcie $(f_L)_t(t_1, t_2, t_3)$ je štatisticky významný, t.j. nemožno ho zanedbať. Medzera medzi elektródami je dôležitý faktor pre proces nitridového leptania.

Má výkon na katóde štatisticky výrazný vplyv na rýchlosť nitridového leptania?

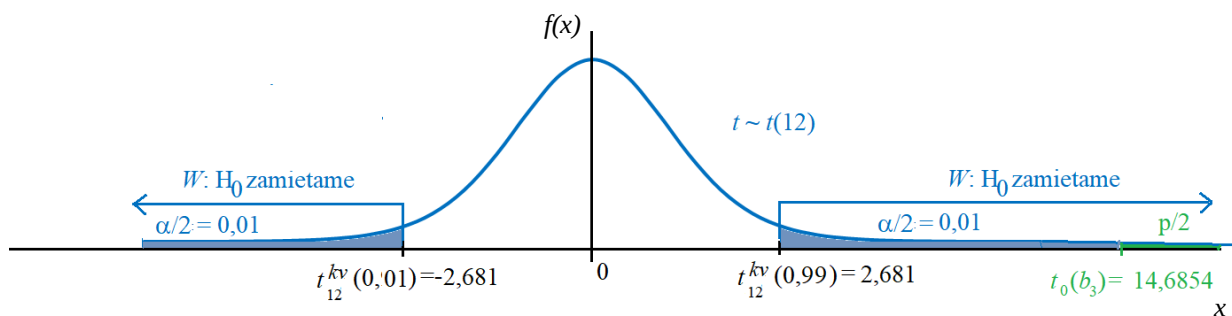
Zvolíme hladinu významnosti napr. 2%. Chceme otestovať, či je samostatné pôsobenie faktora x_3 – výkon na katóde na hodnotu regresnej funkcie lineárneho typu štatisticky významné.

Testujeme hypotézy $H_0 : b_3 = 0 \leq H_1 : b_3 \neq 0$, pretože samostatné pôsobenie kódovaného faktora t_3 vystupuje v transformovanej regresnej funkcii $(f_L)_t(t_1, t_3) = b_0 + b_1 t_1 + b_3 t_3 + b_{13} t_1 t_3$ v predposlednom člene $b_3 t_3$, t.j. budeme testovať nulovosť regresného koeficientu b_3 . Z výstupu Excelu použijeme riadok, v ktorom sa nachádzajú údaje súvisiace s koeficientom b_3 , obr.1.16

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 98,0%	Upper 98,0%
Intercept	776,0625	10,42276905	74,45838	2,29E-17	753,353237	798,771763	748,1190771	804,0059229
t1	-50,8125	10,42276905	-4,87514	0,000382	-73,521763	-28,1032371	-78,7559229	-22,8690771
t3	153,0625	10,42276905	14,6854	4,95E-09	130,353237	175,771763	125,1190771	181,0059229
t1*t3	-76,8125	10,42276905	-7,36968	8,62E-06	-99,521763	-54,1032371	-104,755923	-48,8690771

Obr. 1.16 Tabuľka s bodovými odhadmi koeficientov a im prislúchajúcimi hodnotami

Grafické riešenie je na obr. 1.17



Obr. 1.17 Grafické riešenie pre testovanie hypotézy o koeficiente b_{13}

Kritická oblasť $W = (-\infty; -2,681) \cup (2,681; \infty)$, kde 99% kvantil $t_{12}^{kv}(0,99) = T.INV(0,99; 12) = 2,681$. Platí $|t_0(b_3)| = \frac{153,0625}{10,42276905} = 14,6854 > 2,681 = t_{12}^{kv}(0,99)$.

Z grafického riešenia pre test hypotézy o koeficiente b_3 na obr. 1.17 realizácia testovacej štatistiky $t_0(b_3) = 14,6854 \in W$, preto zamietame $H_0 : b_3 = 0$.

Rovnako p-hodnota $p = 4,95 \cdot 10^{-9} < 0,02 = \alpha$, preto zamietame $H_0 : b_3 = 0$.

P-hodnotu vieme vypočítať vzťahom:

$$p(b_3) = 2 * \min\{F(t_0(b_3)); 1 - F(t_0(b_3))\} = 2 * \text{MIN}(T.DIST(14,6854; 12; 1); 1 - T.DIST(14,6854; 12; 1)) = 4,95 \cdot 10^{-9}.$$

Graficky p-hodnota predstavuje dvojnásobok veľkosti plochy pod krivkou hustoty rozdelenia Studentovho rozdelenia $t(12)$ od realizácie testovacej štatistiky $t_0(b_3)$ vpravo, vid' obr. 1.17.

Výpočet 98% IS pre koeficient b_3 je na obr. 1.18.

	H	I	J	K		H	I	J
79		b3=	153,0625		79		b3=	153,0625
80		SE(b3)=	10,42277		80		SE(b3)=	10,42277
81		t kv 12 (0,99)=	=T.INV(1-0,02/2;12)		81		t kv 12 (0,99)=	2,680998
82		delta=	=J80*J81		82		delta=	27,94342
83		b3-delta=	=J79-J82		83		b3-delta=	125,1191
84		b3+delta=	=J79+J82		84		b3+delta=	181,0059

Obr. 1.18 Výpočet intervalu spoľahlivosti pre koeficient b_3 v Exceli

Z tabuľky na obr. 1.16 je 98% IS pre koeficient b_3 : (125,119; 181,006)



Obr. 1.19 0 nepatrí do 98% IS pre koeficient b_3

Keďže $0 \notin (125,119; 181,006)$, so spoľahlivosťou 98% môžeme hypotézu $H_0 : b_3 = 0$ zamietnuť.

Záver:

Na hladine významnosti $\alpha = 0,02$ možno konštatovať, že regresný koeficient b_3 je štatisticky významný, t.j. samostatný vplyv faktorov t_3 na hodnotu transformovanej regresnej funkcie $(f_L)_t(t_1, t_2, t_3)$ je štatisticky významný, t.j. nemožno ho zanedbať. Výkonom na katóde je dôležitý faktor pre rýchlosť nitrídového leptania.

Ak chceme napr. na hladine významnosti 1% otestovať, či je spoločné pôsobenie faktorov x_1 – medzera medzi elektródami a x_3 – výkon na katóde na skúmanú odozvu štatisticky významné, postavíme hypotézy: $H_0: b_{13} = 0 \leq H_1: b_{13} \neq 0$, pretože spoločné pôsobenie je v regresnom modeli $(f_L)_t(t_1, t_3) = b_0 + b_1 t_1 + b_3 t_3 + b_{13} t_1 t_3$ vyjadrené v poslednom člene $b_{13} t_1 t_3$. Testujeme nulovosť tohto člena regresného modelu pomocou koeficientu b_{13} . Postupujeme rovnako ako v predchádzajúcich prípadoch. Číselné hodnoty potrebné k testovaniu hypotéz berieme z riadku, ktorý predstavuje interakciu kódovaných faktorov t_1 a t_3 , obr. 1.20.

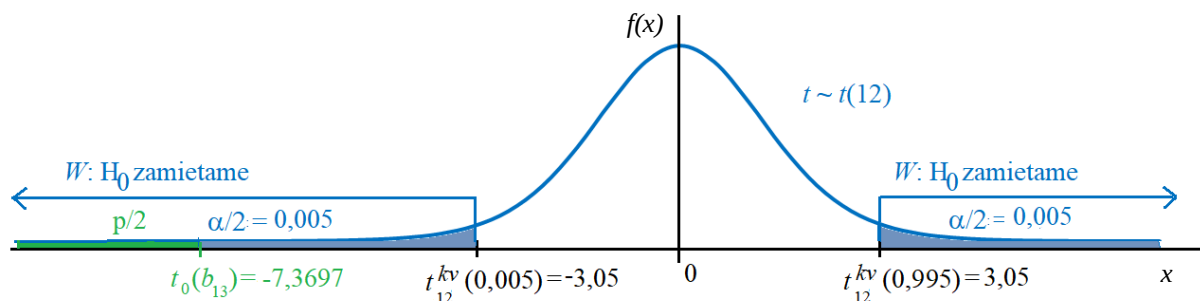
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99,0%	Upper 99,0%
Intercept	776,0625	10,42276905	74,45838	2,29E-17	753,353237	798,771763	744,225739	807,8992607
t1	-50,8125	10,42276905	-4,87514	0,000382	-73,521763	-28,1032371	-82,649261	-18,9757393
t3	153,0625	10,42276905	14,6854	4,95E-09	130,353237	175,771763	121,225739	184,8992607
t1*t3	-76,8125	10,42276905	-7,36968	8,62E-06	-99,521763	-54,1032371	-108,64926	-44,9757393

Obr. 1.20 Tabuľka s bodovými odhadmi koeficientov a im prislúchajúcimi hodnotami

Ak o výsledku testu rozhodujeme pomocou kritickej oblasti $W = (-\infty; -3,05) \cup (3,05; \infty)$, kde 99,5% kvantil Studentovho rozdelenia $t_{12}^{kv}(0,995) = T.INV(0,995; 12) = 3,05$, potom platí $|t_0(b_{13})| = \left| \frac{-76,8125}{10,42276905} \right| = 7,3697 > 3,05 = t_{12}^{kv}(0,995)$. Z grafického riešenia pre test hypotézy o koeficiente b_{13} v tabuľke na obr. 1.20 realizácia testovacej štatistiky $t_0(b_{13}) = -7,3697 \in W$, preto zamietame $H_0: b_{13} = 0$.

Rovnako p-hodnota $p = 8,62 \cdot 10^{-6} < 0,01 = \alpha$, preto zamietame $H_0: b_{13} = 0$.

P-hodnotu vieme vypočítať: $p(b_{13}) = 2 * \min\{F(t_0(b_{13})); 1 - F(t_0(b_{13}))\} = 2 * \min(T.DIST(-7,36968; 12; 1); 1 - T.DIST(-7,36968; 12; 1)) = 8,62 \cdot 10^{-6}$. P-hodnota je rovná dvojnásobku veľkosti plochy pod krivkou hustoty rozdelenia Studentovho rozdelenia $t(12)$ po realizáciu testovacej štatistiky $t_0(b_{13})$, vid' obr. 1.21.



Obr. 1.21 Grafické riešenie pre testovanie hypotézy o koeficiente b_{13}

Výpočet 99% IS pre koeficient b_{13} je na obr. 1.22.

	H	I	J	K
102		b13=	153,0625	
103		SE(b13)=	10,42277	
104		t kv 12 (0,995)=	=T.INV(1-0,01/2;12)	
105		delta=	=J103*J104	
106		b13-delta=	=J102-J105	
107		b13+delta=	=J102+J105	

Obr. 1.22 Výpočet intervalu spoľahlivosti pre koeficient b_{13} v Exceli

Z tabuľky na obr. 1.20 je 99% IS pre koeficient b_{13} : $(-108,6493; -44,9757)$



Obr. 1.23 0 nepatrí do 98% IS pre koeficient b_{13}

Keďže $0 \notin (-108,6493; -44,9757)$, so spoľahlivosťou 99% môžeme hypotézu $H_0 : b_{13} = 0$ zamietnuť.

Záver:

Na hladine významnosti $\alpha = 0,01$ možno konštatovať, že regresný koeficient b_{13} je štatisticky významný, t.j. spoločný vplyv faktorov t_1 a t_3 na hodnotu transformovanej regresnej funkcie $(f_L)_t(t_1, t_2, t_3)$ je štatisticky významný, t.j. nemožno ho zanedbať. Interakcia medzi medzerou medzi elektródami a výkonom na katóde je dôležitá pre rýchlosť nitridového leptania.

Kvalita modelu vyjadrená pomocou R^2 , upraveného R^2_{adj} a prediktívneho R^2_{pred}

Kvalitu získaného regresného modelu môžeme charakterizovať pomocou indexu determinácie (koeficientu determinácie) R^2 . Hodnota R^2 je súčasťou výstupu regresnej analýzy v časti Regresné štatistiky uvedená ako *R Square*, obr. 1.24.

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0,98017906
R Square	0,96075098
Adjusted R Square	0,95093873
Standard Error	41,6910762
Observations	16

Obr. 1.24 Štatistiky R^2 a adjusted R^2 týkajúce sa získaného regresného modelu ako výstup regresnej analýzy v Exceli

Číselnú hodnotu indexu determinácie vieme vypočítať aj z tabuľky ANOVA (obr.1.25), pretože R^2 vyjadruje, aký je podiel variability závislej premennej, ktorý vieme vyjadriť získaným regresným modelom. V našom prípade suma štvorcov odchýlok nameraných hodnôt rýchlosti nitridového leptania od celkového priemeru je SS Total = 531 420,9375. Z toho variabilita vysvetlená regresným modelom je SS Regression = 510 563,1875.

ANOVA					
	df	SS	MS	F	Significance F
Regression	3	510563,1875	170187,7	97,91338	1,05389E-08
Residual	12	20857,75	1738,146		
Total	15	531420,9375			

Obr. 1.25 ANOVA ako súčasť výstupu regresnej analýzy v Exceli

Podiel variability pozorovaných dát, ktorú vieme vysvetliť získaným modelom je $R^2 = \frac{510\,563,1875}{531\,420,9375} = 0,960751$, čo interpretujeme nasledovne: 96,1% variability nameraných hodnôt pri nitridovom leptaní vieme vysvetliť regresným modelom (regresnou funkciou lineárneho typu), iba 3,9% variability nameraných hodnôt zostalo modelom nevysvetlenej.

Hodnotiť kvalitu modelu iba podľa indexu determinácie R^2 nie je správne, pretože pridaním prediktorov (členov v regresnom modeli) R^2 vždy narastie. Väčšiu výpovednú hodnotu má pre nás upravený index determinácie R^2_{adj} uvedený v tabuľke na obr. 1.24 ako *Adjusted R Square*. $R^2_{adj} = 0,95093873$ tiež vyjadruje mieru zhody nameraných dát a fitovaných hodnôt (95,1%), ale zohľadňuje pritom počet členov (prediktorov) v modeli. Upravený index determinácie R^2_{adj} penalizuje model za členy, ktoré sú v ňom navyše, preto po vylúčení nevýznamných faktorov a interakcií R^2_{adj} narastie, viď obr. 1.26. Medzi R^2 a R^2_{adj} by nemal byť „veľký rozdiel“ (v opačnom prípade očakávame prítomnosť nevýznamných členov v modeli).

povodny model (pred vylucenim)		hierarchicky model (po vyluceni)	
Regression Statistics		Regression Statistics	
Multiple R	0,982899	Multiple R	0,9801791
R Square	0,96609	R Square	0,960751
Adjusted R Square	0,936419	Adjusted R Square	0,9509387
Standard Error	47,46117	Standard Error	41,691076
Observations	16	Observations	16

Obr. 1.26 Po odstránení nevýznamných členov z regresného modelu upravený index determinácie narastie

Prediktívne R_{pred}^2 určuje ako presne vie model predikovať správnu hodnotu po odstránení daného údaju zo súboru dát (93,02%). V našom prípade môžeme hovoriť, že náš model vie v rámci experimentálneho priestoru veľmi dobre predikovať dáta. Veľký rozdiel medzi R^2 a R_{pred}^2 naznačuje, že v regresnom modeli je zbytočne veľa členov v pomere k množstvu dát. Možné zdôvodnenie je, že buď sú dáta zlé, alebo boli zvolené nesprávne nezávislé premenné. Excel prediktívnu hodnotu R_{pred}^2 nepočíta. Výstup z Minitabu je na obr. 1.27.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
41,6911	96,08%	95,09%	93,02%

Obr. 1.27 Výstup Minitabu s R^2 , upraveným R^2 a prediktívnym R^2

Kvalita modelu vyjadrená celkovým F testom štatistickej významnosti regresného modelu

Kvalitu modelu vieme vyhodnotiť tiež celkovým F testom štatistickej významnosti regresného modelu, testom adekvátnosti (testom Lack of fit) a analýzou reziduí.

Celkovým F testom štatistickej významnosti nášho regresného modelu $(f_L)_t(t_1, t_3) = b_0 + b_1 t_1 + b_3 t_3 + b_{13} t_1 t_3$ testujeme nulovú hypotézu, že regresný model nie je štatisticky významný, t.j. všetky jeho koeficienty sú nulové oproti alternatívnej hypotéze, že regresný model je štatisticky významný, t.j. aspoň jeden koeficient je nenulový:

$$H_0: b_1 = b_3 = b_{13} = 0 \leq H_1: \text{existuje aspoň jedno } b_i \neq 0$$

Pri celkovom F teste štatistickej významnosti regresného modelu pre testovaciu charakteristiku (štatistiku) platí:

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n - k - 1}} \sim F(k; n - k - 1),$$

kde n je počet nameraných dát ($n = m \cdot 2^k$) a k je počet členov v regresnom modeli. Čítame testovacia štatistika F má Fisherovo rozdelenie so stupňami voľnosti k , $n - k - 1$. Testovacia štatistika porovnáva rozptyl dát vysvetlený regresným modelom s rozptylom, ktorý zostal modelom nevysvetlený. Čím je tento podiel väčší, tým regresný model lepšie vystihuje správanie skúmaných dát. Preto je tento test jednostranný (pravostranný).

Realizáciu testovacej štatistiky, stupne voľnosti, p-hodnotu vieme získať z regresnej analýzy z tabuľky ANOVA na obr. 1.28. Pri výpočte testovacej štatistiky používame hodnoty $SSR = SS$ Regression, a $SSE = SS$ Residual a údaje z nich vypočítané.

ANOVA					
	df	SS	MS	F	Significance F
Regression	3	510563,1875	170187,7	97,91338	1,05389E-08
Residual	12	20857,75	1738,146		
Total	15	531420,9375			

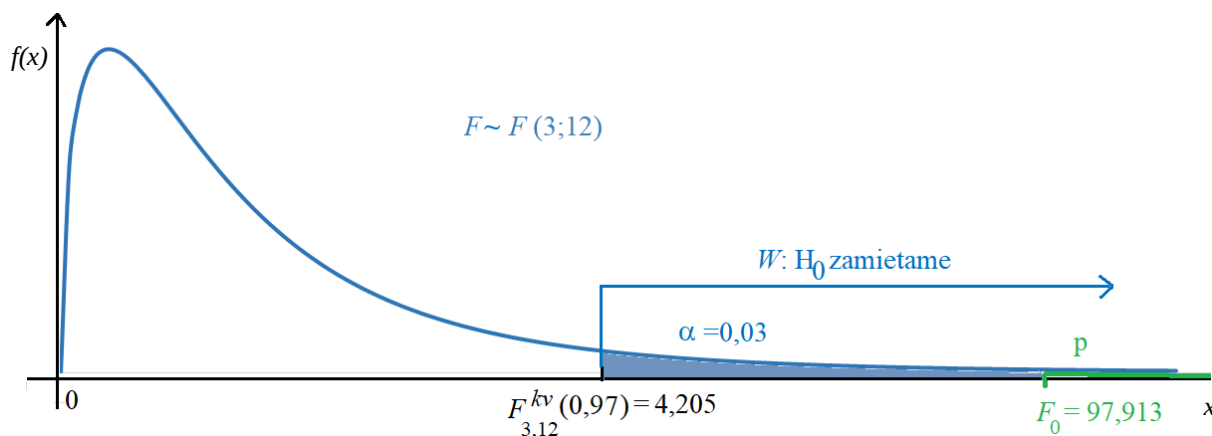
Obr. 1.28 ANOVA ako súčasť výstupu regresnej analýzy v Exceli

Ak zvolíme napr. $\alpha = 0,03$, potom kritická oblasť bude ohraničená kritickou hodnotou $F_{(3,12)}^{kv}(0,97)$. Jej hodnotu vypočítame ako 97% kvantil Fisherovho-Snedecorovho rozdelenia $F.INV(0,97; 3; 12) = 4,205$. Kritická oblasť $W = (4,205; \infty)$.

Realizácia testovacej štatistiky F_0 sa nachádza v tabuľke na obr.1.28 v stĺpci F :

$$F_0 = \frac{\frac{510\,563,1875}{3}}{\frac{20\,857,75}{12}} = \frac{170\,187,7}{1\,738,146} = 97,91338 > 4,205 = F_{(3,12)}^{kv}(0,97)$$

Pretože $F_0 = 97,91338 \in W$ (viď obr. 1.29), zamietame $H_0: b_1 = b_3 = b_{13} = 0$ a prijímame alternatívnu hypotézu H_1 : existuje aspoň jedno $b_i \neq 0$.



Obr. 1.29 Grafické riešenie pre celkový F test o štatistickej významnosti regresného modelu

Pri vyhodnotení podľa p -hodnoty, jej číselnú hodnotu nájdeme v tabuľke na obr.1.28 v stĺpci *Significance F*, alebo vypočítame pomocou distribučnej funkcie v realizácii testovacej štatistiky F_0 : $p = 1 - F(F_0) = 1 - F.DIST(97,91338; 3; 12; 1) = 1,054 \cdot 10^{-8}$

Ak p hodnotu porovnáme so zvolenou hladinou významnosti $\alpha = 0,03$, dostaneme rovnaký záver o tom, že zamietame $H_0: b_1 = b_3 = b_{13} = 0$ a prijímame alternatívnu hypotézu H_1 : existuje aspoň jedno $b_i \neq 0$.

Záver:

Na hladine významnosti $\alpha = 0,03$ možno konštatovať, že regresný model lineárneho typu je štatisticky významný.

F test štatistickej významnosti prínosu k -tej vysvetľujúcej premennej

Súčasťou ANOVY, ktorá obsahuje celkový F test štatistickej významnosti regresného modelu, zvyčajne bývajú aj F testy štatistickej významnosti prínosu k -tých vysvetľujúcich premenných. Rovnako ako pri individuálnych t-testoch štatistickej významnosti regresných koeficientov na základe uvedených testov vieme rozhodnúť, ktorý faktor a interakciu faktorov v regresnom modeli ponechávame ako signifikantné, ktoré členy z modelu je možné vylúčiť. Uvedeným testom testujeme prínos, ktorým dodatočne prispieva k -ta vysvetľujúca premenná k vysvetleniu variability závislej premennej. Význam tohto testu spočíva v tom, že v prípade zamietnutia nulovej hypotézy pri F -teste štatistickej významnosti regresného modelu uvedeného ďalej chceme vedieť, ktorá zo vstupných premenných má štatisticky významný a nevýznamný vplyv na závislú premennú (chceme vedieť, ktoré konkrétne premenné štatisticky významne prispievajú k modelu). Označením dodatočný príspevok k -tej vysvetľujúcej premennej rozumieme prínos tejto premennej za predpokladu, že sa už prv uvažovalo s $(k - 1)$ nezávislými (vysvetľujúcimi) premennými.

Testované hypotézy sú:

H_0 : Dodatočný príspevok k -tej vysvetľujúcej premennej na vysvetlenie variability vysvetľovanej premennej nie je štatisticky významný.

H_1 : Dodatočný príspevok k -tej vysvetľujúcej premennej na vysvetlenie variability vysvetľovanej premennej je štatisticky významný.

Testovacia štatistika:

$$F = \frac{\frac{SSR_{\text{úplný model}} - SSR_{\text{redukovaný model}}}{1}}{\frac{SSE_{\text{úplný model}}}{n - k - 1}} \sim F(1; n - k - 1),$$

kde $SSR_{\text{úplný model}}$ je suma štvorcov odchýlok od celkového priemeru vysvetlená úplným regresným modelom s k prediktormi, $SSR_{\text{redukovaný model}}$ je suma štvorcov odchýlok vysvetlená redukovaným regresným modelom s $k - 1$ prediktormi, ich rozdiel nazývame príspevok k -tej vysvetľujúcej premennej, n je počet nameraných dát a k je počet členov (prediktorov, vysvetľujúcich premenných) v úplnom regresnom modeli. Tento test je pravostranný, nakoľko chceme, aby bol prínos k -tej vysvetľujúcej premennej čo najväčší, v opačnom prípade je možné k -tu vysvetľujúcu premennú z modelu vylúčiť.

Výsledok uvedeného testu významnosti prínosu vieme získať v Minitabe, Excel tento test neposkytuje. Jeho závery týkajúce sa ponechania alebo vylúčenia člena z modelu sú totožné s výsledkom t-testu štatistickej významnosti regresného modelu, napriek tomu, že oba testy sú založené na inom princípe, t.j. testujú niečo iné.

ANOVA z Minitabu na obr. 1.30 poskytuje výsledky F testu štatistickej významnosti prínosu pre nami zvolené (ponechané) premenné. Ukážka výpočtu hodnoty príspevku cez rozdiel SSR pre model s k -tou vysvetľujúcou premennou a bez nej pre transformovaný faktor t_1 - medzera medzi elektródami v Exceli je na obr. 1.31.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	3	510563	170188	97,91	0,000
Linear	2	416161	208080	119,71	0,000
medzera	1	41311	41311	23,77	0,000
vykon na katode	1	374850	374850	215,66	0,000
2-Way Interactions	1	94403	94403	54,31	0,000
medzera*vykon na katode	1	94403	94403	54,31	0,000
Error	12	20858	1738		
Lack-of-Fit	4	2837	709	0,31	0,860
Pure Error	8	18021	2253		
Total	15	531421			

Obr. 1.30 Výstup Minitabu pre F test o štatistickej významnosti prínosu faktora medzera

Chceme za predpokladu, že sa už v predchádzajúcej časti uvažovalo o ostatných členov modelu napr. na hladine významnosti 0,1 overiť štatistickú významnosť prínosu premennej medzera medzi elektródami na vysvetlenie variability rýchlosti nitridového leptania.

Testované hypotézy pre F test o štatistickej významnosti prínosu faktora medzera medzi elektródami sú:

H_0 : Dodatočný príspevok faktora medzera na vysvetlenie variability rýchlosti nitridového leptania nie je štatisticky významný.

H_1 : Dodatočný príspevok faktora medzera na vysvetlenie variability rýchlosti nitridového leptania je štatisticky významný.

	A	B	C	D	E	F	G	H	I	J
1							t1:			
2	t1	t3	t1*t3	yi			y predik s t1	y predik bez t1	(delta s t1)^2	(delta bez t1)^2
3	-1	-1	=A3*B3	550			=G\$25+G\$26*A3+G\$27*B3+G\$28*C3	=H\$25+H\$26*B3+H\$27*C3	=(G3-G\$25)^2	=(H3-G\$25)^2
4	-1	-1	=A4*B4	604			=G\$25+G\$26*A4+G\$27*B4+G\$28*C4	=H\$25+H\$26*B4+H\$27*C4	=(G4-G\$25)^2	=(H4-G\$25)^2
5	1	-1	=A5*B5	669			=G\$25+G\$26*A5+G\$27*B5+G\$28*C5	=H\$25+H\$26*B5+H\$27*C5	=(G5-G\$25)^2	=(H5-G\$25)^2
6	1	-1	=A6*B6	650			=G\$25+G\$26*A6+G\$27*B6+G\$28*C6	=H\$25+H\$26*B6+H\$27*C6	=(G6-G\$25)^2	=(H6-G\$25)^2
7	-1	-1	=A7*B7	633			=G\$25+G\$26*A7+G\$27*B7+G\$28*C7	=H\$25+H\$26*B7+H\$27*C7	=(G7-G\$25)^2	=(H7-G\$25)^2
8	-1	-1	=A8*B8	601			=G\$25+G\$26*A8+G\$27*B8+G\$28*C8	=H\$25+H\$26*B8+H\$27*C8	=(G8-G\$25)^2	=(H8-G\$25)^2
9	1	-1	=A9*B9	642			=G\$25+G\$26*A9+G\$27*B9+G\$28*C9	=H\$25+H\$26*B9+H\$27*C9	=(G9-G\$25)^2	=(H9-G\$25)^2
10	1	-1	=A10*B10	635			=G\$25+G\$26*A10+G\$27*B10+G\$28*C10	=H\$25+H\$26*B10+H\$27*C10	=(G10-G\$25)^2	=(H10-G\$25)^2
11	-1	1	=A11*B11	1037			=G\$25+G\$26*A11+G\$27*B11+G\$28*C11	=H\$25+H\$26*B11+H\$27*C11	=(G11-G\$25)^2	=(H11-G\$25)^2
12	-1	1	=A12*B12	1052			=G\$25+G\$26*A12+G\$27*B12+G\$28*C12	=H\$25+H\$26*B12+H\$27*C12	=(G12-G\$25)^2	=(H12-G\$25)^2
13	1	1	=A13*B13	749			=G\$25+G\$26*A13+G\$27*B13+G\$28*C13	=H\$25+H\$26*B13+H\$27*C13	=(G13-G\$25)^2	=(H13-G\$25)^2
14	1	1	=A14*B14	868			=G\$25+G\$26*A14+G\$27*B14+G\$28*C14	=H\$25+H\$26*B14+H\$27*C14	=(G14-G\$25)^2	=(H14-G\$25)^2
15	-1	1	=A15*B15	1075			=G\$25+G\$26*A15+G\$27*B15+G\$28*C15	=H\$25+H\$26*B15+H\$27*C15	=(G15-G\$25)^2	=(H15-G\$25)^2
16	-1	1	=A16*B16	1063			=G\$25+G\$26*A16+G\$27*B16+G\$28*C16	=H\$25+H\$26*B16+H\$27*C16	=(G16-G\$25)^2	=(H16-G\$25)^2
17	1	1	=A17*B17	729			=G\$25+G\$26*A17+G\$27*B17+G\$28*C17	=H\$25+H\$26*B17+H\$27*C17	=(G17-G\$25)^2	=(H17-G\$25)^2
18	1	1	=A18*B18	860			=G\$25+G\$26*A18+G\$27*B18+G\$28*C18	=H\$25+H\$26*B18+H\$27*C18	=(G18-G\$25)^2	=(H18-G\$25)^2
19									SS: =SUM(I3:I18)	=SUM(J3:J18)
20										
21										
22										
23										
24										
25										
26										
27										
28										

	A	B	C	D	E	F	G	H	I	J
1	po vyluceni statisticky nevyznamnych faktorov:						t1:			
2	t1	t3	t1*t3	yi			y predik s t1	y predik bez t1	(delta s t1)^2	(delta bez t1)^2
3	-1	-1	1	550			597	546,1875	32063,37891	52842,51563
4	-1	-1	1	604			597	546,1875	32063,37891	52842,51563
5	1	-1	-1	669			649	699,8125	16144,87891	5814,0625
6	1	-1	-1	650			649	699,8125	16144,87891	5814,0625
7	-1	-1	1	633			597	546,1875	32063,37891	52842,51563
8	-1	-1	1	601			597	546,1875	32063,37891	52842,51563
9	1	-1	-1	642			649	699,8125	16144,87891	5814,0625
10	1	-1	-1	635			649	699,8125	16144,87891	5814,0625
11	-1	1	-1	1037			1056,75	1005,9375	78785,47266	52842,51563
12	-1	1	-1	1052			1056,75	1005,9375	78785,47266	52842,51563
13	1	1	1	749			801,5	852,3125	647,0664063	5814,0625
14	1	1	1	868			801,5	852,3125	647,0664063	5814,0625
15	-1	1	-1	1075			1056,75	1005,9375	78785,47266	52842,51563
16	-1	1	-1	1063			1056,75	1005,9375	78785,47266	52842,51563
17	1	1	1	729			801,5	852,3125	647,0664063	5814,0625
18	1	1	1	860			801,5	852,3125	647,0664063	5814,0625
19									SS: 510563,1875	469252,625
20										
21										

Obr. 1.31 Výpočet prínosu t_1 pre F test o štatistickej významnosti prínosu t_1

Pre výpočet hodnoty testovacej štatistiky je potrebné dopočítať sumu reziduálnych štvorcov pre úplný model, t.j. s k vysvetľujúcimi premennými:

$SSE_{\text{úplný model}} = SS_{\text{Total}} - SSR_{\text{úplný model}}$, kde SS_{Total} je celková suma štvorcov odchýlok od priemeru, $SSR_{\text{úplný model}}$ je suma štvorcov odchýlok od celkového priemeru vysvetlená úplným regresným modelom s k prediktormi. Výpočet týchto hodnôt v Exceli je na obr. 1.32.

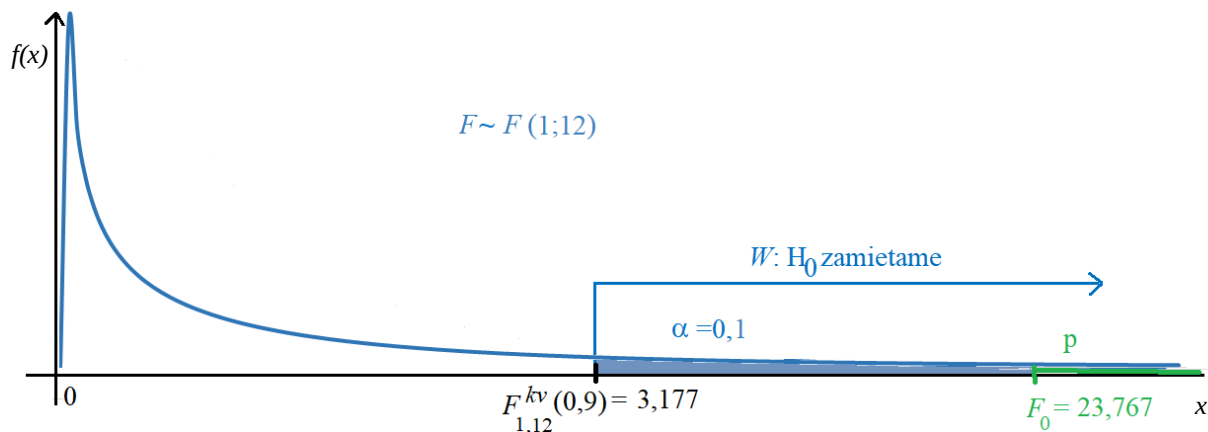
	A	B	C		A	B	C	D	E	F
1	ANOVA				1	ANOVA				
2		df	SS		2		df	SS	MS	F
3	Regression	3	510563,1875		3	Regression	3	510563,2	170187,7	97,91338
4	Residual	12	20857,75		4	Residual	12	20857,75	1738,146	1,05389E-08
5	Total	15	531420,9375		5	Total	15	531420,9		
6					6					
7					7					
8	prinos t1:	41310,5625			8	prinos t1:	41310,56			
9	SSR pre t1:	510563,1875			9	SSR pre t1:	510563,2			
10	SSE(t1):	=C5-B9			10	SSE(t1):	20857,75			
11	F0=	=(B8/1)/(B10/(16-3-1))			11	F0=	23,76703			

Obr. 1.32 Výpočet SSE pre úplný model pre F test o štatistickej významnosti prínosu t_1

Realizácia testovacej štatistiky $F_0(t_1) = \frac{41\,310,56}{\frac{1}{20\,857,75}} = 23,76703$.

Kritická oblasť bude pre $\alpha = 0,1$ ohraničená kritickou hodnotou $F_{(1,12)}^{kv}(0,9)$. Jej hodnotu vypočítame ako 90% kvantil Fisherovho rozdelenia, t.j. $F.INV(0,9; 1; 12) = 3,176549$. Kritická oblasť je potom $W = (3,177; \infty)$.

Pretože vypočítaná hodnota testovacej štatistiky $F_0(t_1) = 23,767$ leží v kritickej oblasti W (viď obr. 1.33), zamietame nulovú hypotézu H_0 a prijímame alternatívnu hypotézu H_1 .



Obr. 1.33 Grafické riešenie pre F test o štatistickej významnosti prínosu t_1

Pri vyhodnotení podľa p -hodnoty: $p = 1 - F(F_0) = 1 - F.DIST(23,767; 1; 12; 1) = 0,000382$. Vo výstupe Minitabu v časti ANOVA sa nachádza P-Value rovná 0,000 nakoľko daný softvér uvádza výsledky iba v tvare desatinného čísla s maximálne tromi platnými číslicami. Ak p -hodnotu porovnáme so zvolenou hladinou významnosti $\alpha = 0,1$, dostaneme rovnaký záver, t.j. zamietame H_0 a prijímame alternatívnu hypotézu H_1 .

Záver:

Na hladine významnosti $\alpha = 0,1$ za predpokladu, že sa už uvažovalo s premennou t_3 -výkon na katóde a interakciou t_1 a t_3 , t.j. vzájomným pôsobením faktorov medzera na elektróde a výkon na katóde, možno považovať dodatočný prínos premennej t_1 - medzera medzi elektródami štatisticky významný.

Obdobným spôsobom je možné testovať dodatočný prínos všetkých členov modelu, t.j. všetkých prediktorov.

Kvalita regresného modelu vyjadrená testom dostatočného prispôsobenie regresného modelu (testom Lack of fit)

Pri teste Lack of fit testujeme adekvátnosť (dostatočnosť) regresného modelu. Vzhľadom na to, že merania boli robené iba na dvoch úrovniach, je potrebné doplniť opakované merania (zvyčajne 3-5) v centrálnom bode so súradnicami [1cm; 162,5 cm³/min; 300 W]. Nech boli namerané tieto hodnoty: 779; 770; 775; 777; 760 [10⁻¹⁰ m/min].

Nech zvolená hladina významnosti je $\alpha = 0,05$.

Testované hypotézy pri tomto teste:

Nulová hypotéza je $H_0: SSLF = SSPE$. Možný je tiež zápis: No LOF – regresný model nie je nedostatočný.

Alternatívna hypotéza $H_1: SSLF > SSPE$, resp. iný zápis: LOF – regresný model je nedostatočný.

Testovacia štatistika:

$$F = \frac{\frac{SSLF}{df_{LF}}}{\frac{SSPE}{df_{PE}}} = \frac{MSLF}{MSPE} \sim F(df_{LF}; df_{PE}),$$

Pri tomto druhu testu reziduá, resp. sumu štvorcov reziduálnych odchýlok SSE rozdeľujeme na také, ktoré vznikli chybou merania, t.j. v dôsledku čistej chyby (pure error) a také, ktoré vznikli z dôvodu nedostatočného prispôsobenia regresného modelu (lack of fit). V tomto prípade bude model tým lepšie vystihovať správanie dát, čím bude suma štvorcov odchýlok spôsobená nedostatočnosťou modelu čo najmenšia.

Excel výpočet testu Lack of fit neposkytuje. Ukážka výpočtu veľkosti čistej chyby PE napr. pre bod [-1;-1] a prislúchajúce stupne voľnosti df je na obr. 1.34. Stupne voľnosti určíme ako počet meraní v danom bode zmenšený o jedna.

	A	B	C	D	E	F
1	po vyluceni statisticky nevyznamnych faktorov:					
2	t1	t3	t1*t3	yi	PE	df pre PE
3	-1	-1	1	550	=DEVSQ(D3:D4;D7:D8)	3
4	-1	-1	1	604		
5	1	-1	-1	669	=DEVSQ(D5:D6;D9:D10)	3
6	1	-1	-1	650		
7	-1	-1	1	633		
8	-1	-1	1	601		
9	1	-1	-1	642		
10	1	-1	-1	635		
11	-1	1	-1	1037	=DEVSQ(D11:D12;D15:D16)	3
12	-1	1	-1	1052		
13	1	1	1	749	=DEVSQ(D13:D14;D17:D18)	3
14	1	1	1	868		
15	-1	1	-1	1075		
16	-1	1	-1	1063		
17	1	1	1	729		
18	1	1	1	860		
19	0	0	0	779	=DEVSQ(D19:D23)	4
20	0	0	0	770		
21	0	0	0	775		
22	0	0	0	777		
23	0	0	0	760		
24					=SUM(E3:E23)	16

Máme 4 namerané hodnoty v bode [-1;-1], preto stupne voľnosti df pre tento bod sú rovné $4-1=3$.

Čistá chyba PE v bode [-1;-1] je vyjadrená ako suma štvorcov odchýlok nameraných hodnôt y od priemernej nameranej hodnoty 597.

Súčet PE a df pre všetkých 5 experimentálnych bodov dosádzame do tabuľky ANOVA.

Obr. 1.34 Výpočet čistej chyby a výpočet stupňov voľnosti pre čistú chybu

Tabuľku ANOVA v Exceli doplníme o dva riadky, ktoré predstavujú rozklad sumy štvorca reziduálnych odchýlok $SSE = SSLF + SSPE$.

	A	B	C	D	E	F	G
145		df	SS	MS	F	F krit	Significance F
146	Regression	3	510563,1875	170187,729166667	136,82378175	=F.INV(0,95;B146;B147)	4,20574614802358E-12
147	Residual	17	21145,3839285714	1243,84611344538			
148	Lack of fit	=B147-B149	=C147-C149	=C148/B148	=D148/D149	=F.INV(0,95;B148;B149)	=1-F.DIST(E148;B148;B149;1)
149	Pure error	=F24	=E24	=C149/B149			
150	Total	20	531708,571428571				

	A	B	C	D	E	F	G
142	$\alpha = 0,05$						
143	test Lack of fit (LOF)						
144	ANOVA						
145		df	SS	MS	F	F krit	Significance F
146	Regression	3	510563,188	170187,7	136,8238	3,196777	4,20575E-12
147	Residual	17	21145,3839	1243,846			
148	Lack of fit	1	56,8339286	56,83393	0,04312	4,493998	0,838119093
149	Pure error	16	21088,55	1318,034			
150	Total	20	531708,571				

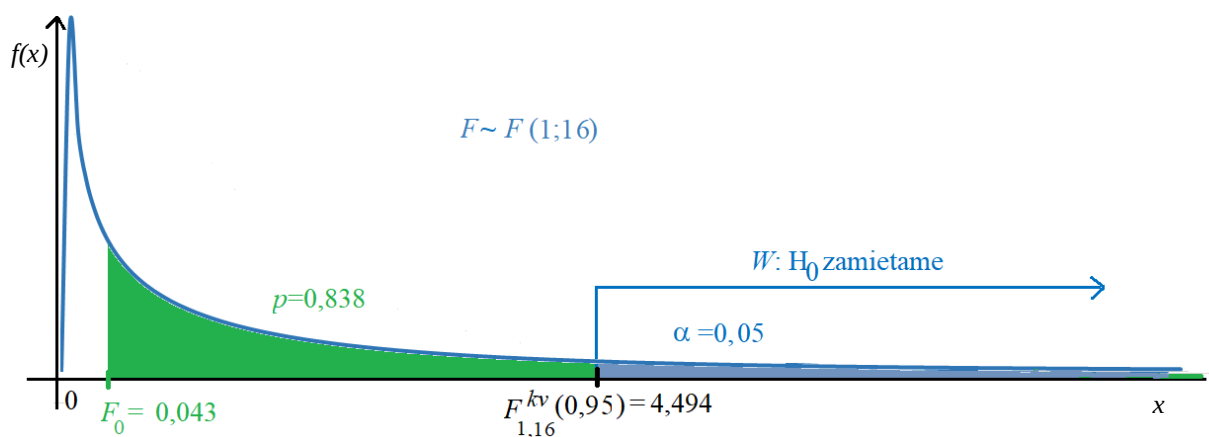
Obr. 1.35 Výpočet hodnôt potrebných pri teste Lack of fit

Testovacia štatistika má Fisherovo rozdelenie s 1 a 16 stupňami voľnosti, pretože df_{PE} sme vypočítali rovné 16 a odtiaľ $df_{LF} = 17 - 16 = 1$.

$$\text{Realizácia testovacej štatistiky } F_0 = \frac{\frac{56,834}{\frac{1}{21\,088,55}}}{16} = \frac{56,834}{1\,318,034} = 0,04312$$

Test je jednostranný, preto kritická oblasť bude pre $\alpha = 0,05$ ohraničená kritickou hodnotou $F_{(1,16)}^{kv}(0,95)$. Jej hodnotu vypočítame ako 95% kvantil Fisherovho rozdelenia, t.j. $F.INV(0,95; 1; 16) = 4,493998$. Kritická oblasť je potom $W = (4,494; \infty)$.

Pretože vypočítaná hodnota testovacej štatistiky $F_0 = 0,043$ neleží v kritickej oblasti W (viď obr. 1.36), nezamietame, t.j. prijímame nulovú hypotézu H_0 .



Obr. 1.36 Grafické riešenie pre test o adekvátnosti regresného modelu

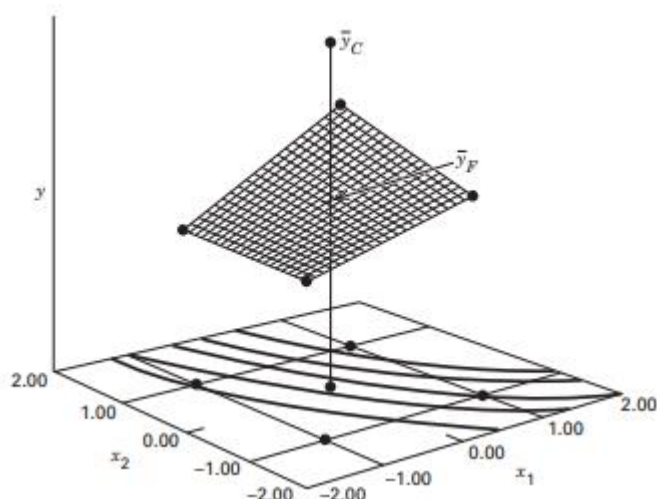
Pri vyhodnotení podľa p -hodnoty: $p = 1 - F(F_0) = 1 - F.DIST(0,043; 1; 16; 1) = 0,838119093$ túto hodnotu porovnávame so zvolenou hladinou významnosti $\alpha = 0,05$, pričom musíme dostať rovnaký záver, t.j. že nezamietame H_0 a zamietame alternatívnu hypotézu H_1 .

Záver:

Na hladine významnosti $\alpha = 0,05$ možno konštatovať, že regresný model lineárneho typu nie je nedostatočný (je dostatočný) a nie je potrebné do modelu zaraďovať kvadratické členy.

Test krivosti

Na určenie toho, či je potrebné do modelu zaradiť kvadratické členy sa používa aj test krivosti, ktorý porovnáva priemer odozvy vo faktoriálnych bodoch s nameraným priemerom odozvy v centrálnom bode, obr. 1.37.



Obr. 1.37 Experiment 2^2 s pridaným centrálnym bodom [Montgomery]

Všeobecný zápis úplného kvadratického modelu s k faktormi vyjadrenými v transformovaných jednotkách je $(\hat{f}_L)_t(t_1, t_2, t_3) = b_0 + \sum_{j=1}^k b_j t_j + \sum \sum_{i < j} b_{ij} t_i t_j + \sum_{j=1}^k b_{jj} t_j^2$. Chceme zistiť, či sú v regresnom modeli kvadratické členy signifikantné. Testujeme hypotézy:

Nulová hypotéza hovorí, že kvadratické členy v regresnom modeli nie sú potrebné, čo zapíšeme $H_0: \sum_{j=1}^k b_{jj} = 0$, alternatívna hypotéza tvrdí opak, t.j. $H_1: \sum_{j=1}^k b_{jj} \neq 0$.

Test je pravostranný.

Testovacia štatistika:

$$F = \frac{\frac{SSCPQ}{1}}{\frac{SSPE}{df_{PE}}} \sim F(1; df_{PE}), \text{ kde } SSCPQ = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C} \text{ je suma štvorcov čistého kvadratického}$$

zakrivenia a n_F je počet meraní vo faktoriálnych bodoch, n_C je počet replikácií v centrálnom bode, y_i sú namerané hodnoty v centrálnom bode, $\bar{y}_F$ je priemerná odozva vo faktoriálnych bodoch a $\bar{y}_C$ je priemerná odozva v centrálnom bode. Výpočet hodnoty testovacej štatistiky realizovaný v Exceli je na obr. 1.38. Čistú chybu PE bolo potrebné po pridaní nameraných hodnôt odozvy v centrálnom bode CP opätovne prepočítať, vid' obr.1.39.

	A	B	C	D	E		A	B	C	D	E
1	central point:	779		factorial point:	1052	1	central point:	779		factorial point:	1052
2		770			550	2		770			550
3		775			601	3		775			601
4		777			635	4		777			635
5		760			669	5		760			669
6	y priem=	=AVERAGE(B1:B5)			633	6	y priem=	772,2			633
7	nC=	5			650	7	nC=	5			650
8					604	8					604
9					642	9					642
10	SSCPQ=	=((E18*B7*(E17-B6)^2))/(E18+B7)			1037	10	SSCPQ=	56,83393			1037
11	F0=	=B10/(21088,55/16)			749	11	F0=	0,04312			749
12					860	12					860
13					868	13					868
14					1075	14					1075
15					1063	15					1063
16					729	16					729
17				y priem=	=AVERAGE(E1:E16)	17				y priem=	776,0625
18				nF=	16	18				nF=	16

Obr. 1.38 Výpočet F_0 pri teste krivosti realizovaný v Exceli

	E	F	G	H	I		E	F	G	H	I
1	A	C	y	SS PE	df	1	A	C	y	SS PE	df
2	-1	-1	550	=DEVSQ(G2;G3;G6;G7)	3	2	-1	-1	550	3570	3
3	-1	-1	604			3	-1	-1	604		
4	1	-1	669	=DEVSQ(G4;G5;G8;G9)	3	4	1	-1	669	646	3
5	1	-1	650			5	1	-1	650		
6	-1	-1	633			6	-1	-1	633		
7	-1	-1	601			7	-1	-1	601		
8	1	-1	642			8	1	-1	642		
9	1	-1	635			9	1	-1	635		
10	-1	1	1037	=DEVSQ(G10;G11;G14;G15)	3	10	-1	1	1037	784,75	3
11	-1	1	1052			11	-1	1	1052		
12	1	1	749	=DEVSQ(G12;G13;G16;G17)	3	12	1	1	749	15857	3
13	1	1	868			13	1	1	868		
14	-1	1	1075			14	-1	1	1075		
15	-1	1	1063			15	-1	1	1063		
16	1	1	729			16	1	1	729		
17	1	1	860			17	1	1	860		
18	0	0	779	=DEVSQ(G18;G22)	4	18	0	0	779	230,8	4
19	0	0	770			19	0	0	770		
20	0	0	775			20	0	0	775		
21	0	0	777			21	0	0	777		
22	0	0	760			22	0	0	760		
23				=SUM(H2:H22)	=SUM(I2:I22)	23				21088,55	16

Obr. 1.39 Výpočet čistej chyby PE pri teste krivosti

Test krivosti ako súčasť ANOVY v Minitabe je možné vidieť na obr. 1.40.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	4	510620	127655	96,85	0,000
Linear	2	416161	208080	157,87	0,000
A	1	41311	41311	31,34	0,000
C	1	374850	374850	284,40	0,000
2-Way Interactions	1	94403	94403	71,62	0,000
A*C	1	94403	94403	71,62	0,000
Curvature	1	57	57	0,04	0,838
Error	16	21089	1318		
Total	20	531709			

Obr. 1.40 Teste krivosti realizovaný v Minitabe

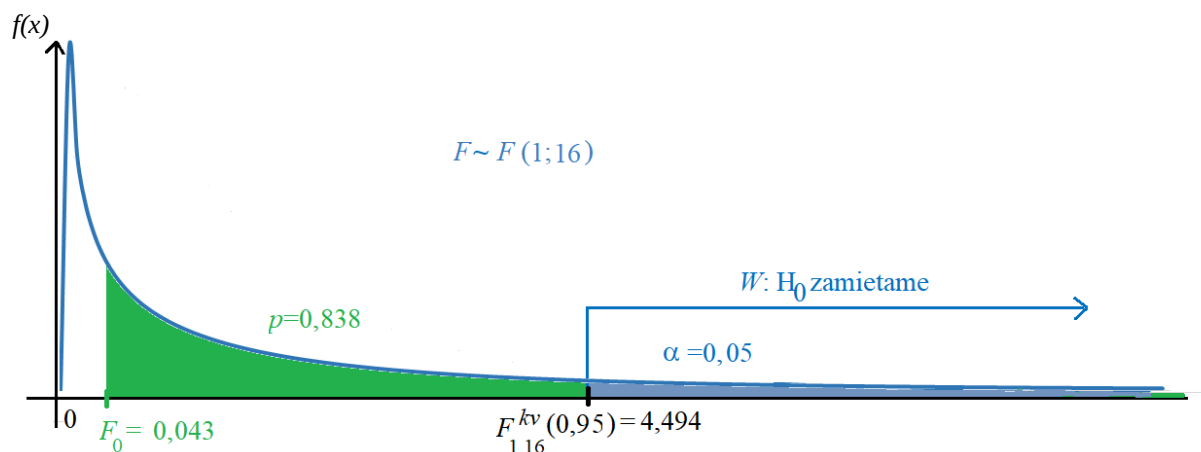
Realizácia testovacej štatistiky $F_0 = \frac{\frac{56,834}{1}}{\frac{21088,55}{16}} = 0,04312$.

Ak zvolíme $\alpha = 0,05$, potom je potrebné vypočítať $F_{1,16}^{kv}(0,95)$, pretože test je jednostranný. Túto hodnotu vypočítame ako 95% kvantil Fisherovho rozdelenia $F(1;16)$, t.j. $F.INV(0,95;1;16) = 4,493998$. Kritická oblasť je $W = (4,494; \infty)$. Pretože vypočítaná hodnota testovacej štatistiky $F_0 = 0,043$ neleží v kritickej oblasti W (viď obr. 1.41), nezamietame, t.j. prijímame nulovú hypotézu $H_0: \sum_{j=1}^k b_{jj} = 0$.

Pri vyhodnocovaní testu hypotéz podľa p -hodnoty:

$$p = 1 - F(F_0) = 1 - F.DIST(0,04312; 1; 16; 1) = 0,838119495.$$

Túto hodnotu porovnávame so zvolenou hladinou významnosti $\alpha = 0,05$, pričom musíme dostať rovnaký záver, t.j. že nezamietame $H_0: \sum_{j=1}^k b_{jj} = 0$ a zamietame alternatívnu hypotézu $H_1: \sum_{j=1}^k b_{jj} \neq 0$.



Obr. 1.41 Grafické riešenie pre test krivosti

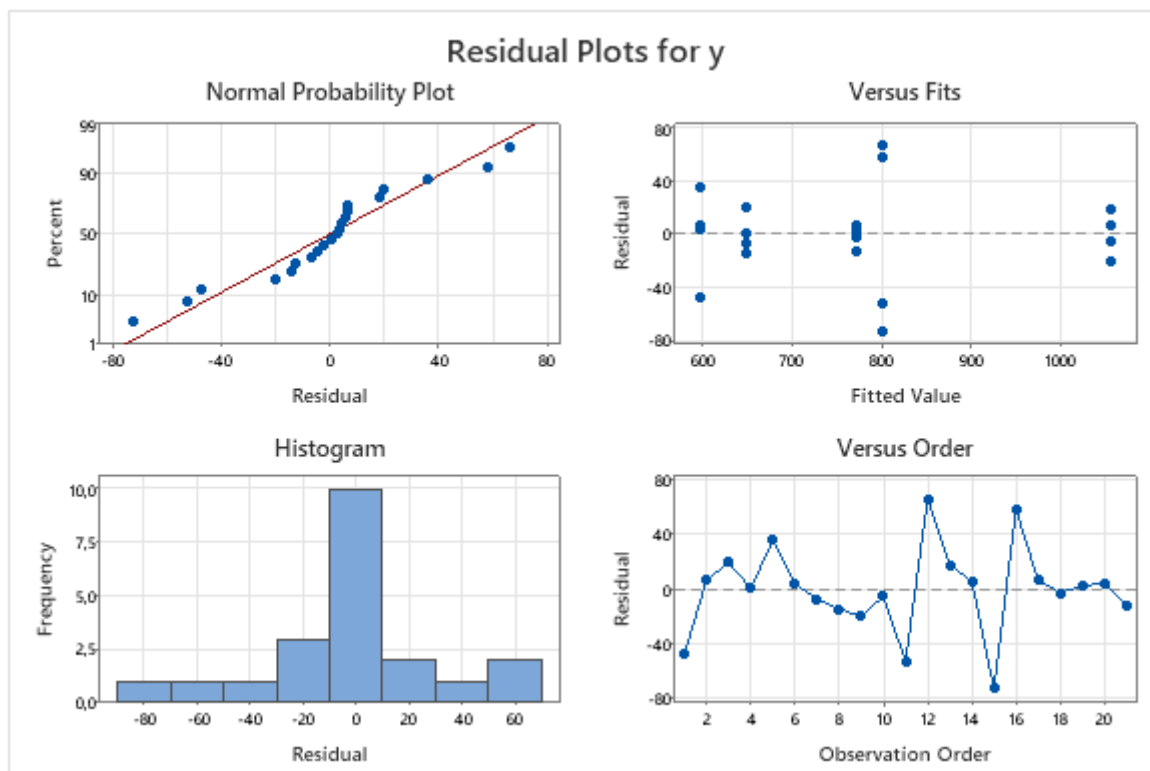
Záver:

Na hladine významnosti $\alpha = 0,05$ prijímame hypotézu, že do regresného modelu nie je potrebné zaraďovať kvadratické členy, t.j. nie je potrebný úplný kvadratický regresný model.

x

Grafická analýza reziduí regresného modelu

Súčasťou posúdenia kvality akéhokoľvek regresného modelu by mala byť minimálne grafická analýza reziduí, kedy vizuálne kontrolujeme normalitu a nezávislosť reziduí. Výstup Minitabu s grafmi reziduí je vidieť na obr. 1.42.



Obr. 1.42 Výstup Minitabu umožňujúci analýzu reziduí

Vizuálnu kontrolu normálneho rozdelenia reziduí umožňuje normálny pravdepodobnostný graf (Normal Probability Plot) a histogram (Histogram). Normálny pravdepodobnostný graf zobrazuje rozdelenie reziduí oproti teoretickým kvantilom normálneho rozdelenia. Kvantily, t.j. percentily predstavujú podiel dát menších ako dané reziduum, ak by rezidá mali normálne rozdelenie. V ideálnom prípade by mali zobrazené body ležať na červenej diagonálnej priamke, čo interpretujeme ako splnenú normalitu analyzovaných dát. V tomto prípade sa reziduá väčšinou nachádzajú blízko čiary, sú však viditeľné určité odchýlky v dolnej a hornej časti grafu, čo naznačuje mierne odchýlky od normálneho rozdelenia. Tieto odchýlky môžu naznačovať prítomnosť odľahlých hodnôt alebo systematických chýb. Ak na posúdenie normality reziduí použijeme histogram, tento by mal mať pre normálne rozdelené reziduá, približne tvar zvona. Náš histogram ukazuje veľmi mierne asymetrické rozdelenie, čo môže naznačovať mierne skreslenie dát.

Graf reziduí verzus fitované hodnoty (Residuals versus Fits) ukazuje, či existuje nejaká systematická závislosť medzi reziduami a predikovanými hodnotami. Ideálne by reziduá mali byť rozptýlené náhodne okolo nulovej osi dodržiavajúc zákon troch sigma. V našom prípade je viditeľný relatívne náhodný rozptyl reziduí, ale je vidieť rôzne zhluky pre rôzne fitované hodnoty, čo môže naznačovať heteroskedasticitu (nekonzistentnosť rozptylu) reziduí.

Z grafu reziduí verzus poradie (Residuals vs Order) vieme zistiť, či existuje trend, t.j. nejaká funkčná závislosť v zaznamenanom časovom poradí reziduí. Ideálne by mali byť hodnoty reziduí náhodne rozptýlené, čo by naznačovalo absenciu autokorelácie. Správanie reziduí v čase v našom prípade nenaznačuje žiadny trend. Reziduá sú náhodne rozptýlené bez vzoru v poradí, čo naznačuje, že nie je prítomná autokorelácia.

Celkovo možno zhodnotiť, že grafy naznačujú, že model je vhodný a reziduá majú požadované vlastnosti pre „dobré sediace“ model.

Test normality

Testy normality sú dôležité najmä preto, že mnohé štatistické testy a metódy predpokladajú splnenú podmienku normality dát. Existuje niekoľko rôznych testov normality, napr. test nulovej šikmosti, test nulovej špicatosti, kombinovaný test, Kolmogorov-Smirnov test a mnohé ďalšie. Každý z týchto testov má svoje špecifické podmienky a použitie. Cieľom testu normality v našom prípade je overiť, či reziduá po fitovaní dát regresným modelom sledujú normálne rozdelenie, čo je dôležitý predpoklad pri analýze výsledkov regresného modelu pomocou ANOVY. V Minitabe sú k dispozícii, Ryan-Joiner test (modifikovaný Shapiro-Wilk test), Anderson-Darling test a Kolmogorov-Smirnovovho test:

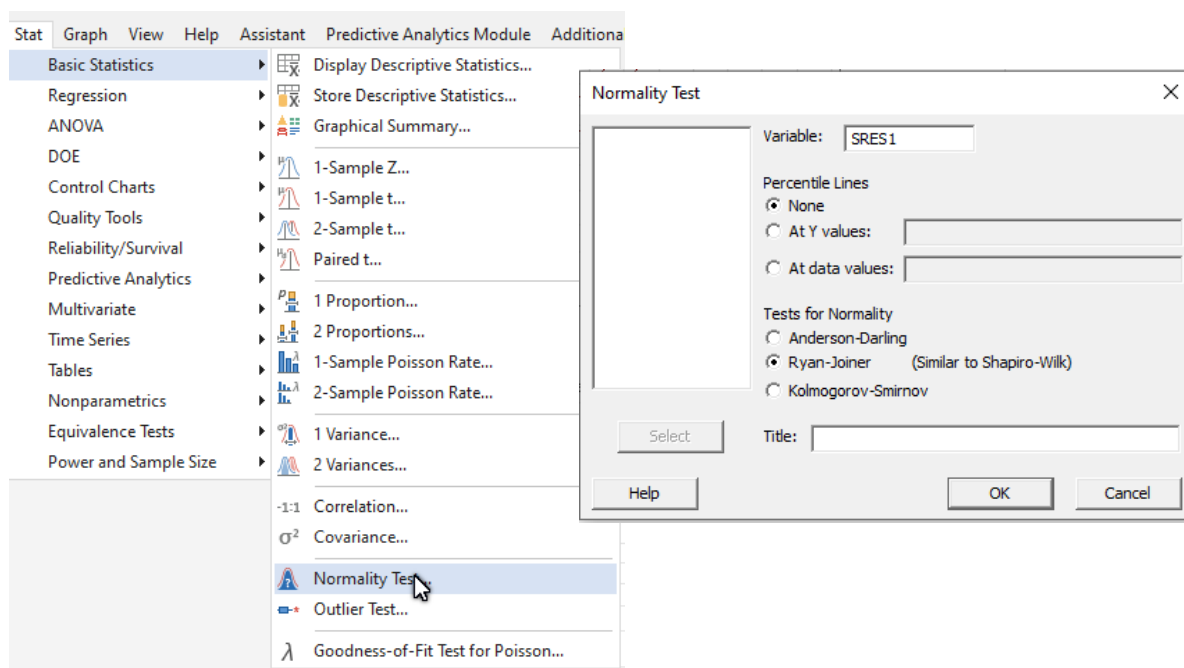
Shapiro-Wilk test (S-W test) je odporúčaný pre malé vzorky - zvyčajne do 50 údajov, pretože je citlivý na malé odchýlky od normality. Pre faktorový dizajn s replikáciami je Shapiro-Wilk test najčastejšie používaný vo väčšine štatistických programov (napríklad R, Python, SPSS atď.). Minitab poskytuje jeho modifikovanú verziu, tzv. **Ryan-Joiner test (R-J test)**. Shapiro-Wilk test (S-W test) sa zakladá na pomere variability, zatiaľ čo Ryan-Joiner test (R-J test) používa koreláciu medzi usporiadanými dátami a očakávanými kvantilmi.

Anderson-Darling test (A-D test) je tiež často používaný, najmä keď je dôležité zohľadniť extrémne hodnoty (outliers), pretože je citlivejší na odchýlky v chvostoch rozdelenia. Je považovaný za robustnejší ako Kolmogorov-Smirnov test (K-S test).

- Kolmogorov-Smirnov test (K-S test) môže byť použitý na zistenie zhody s normálnym rozdelením, avšak pre malé vzorky býva menej presný ako Shapiro-Wilk test.

Pri väčších vzorkách sú preferované Kolmogorov-Smirnov alebo Anderson-Darling, pretože poskytujú robustnejšie výsledky. Údaje by však nemali obsahovať extrémne odľahlé hodnoty (výrazne väčšie než očakávané rozmedzie 2 alebo 3 štandardné odchýlky od priemeru pre štandardizované reziduá), pretože môžu výrazne ovplyvniť výsledky testu. Ak sú prítomné odľahlé hodnoty (Grubbsov a Dixonov test), je vhodné ich najskôr identifikovať a zvážiť ich vplyv.

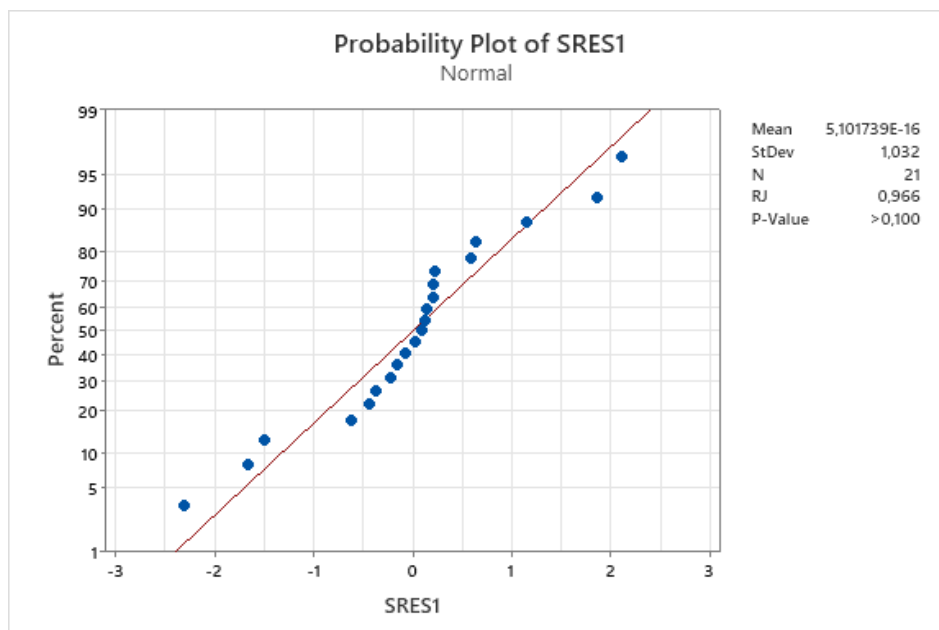
Pred testom normality reziduí je potrebné hodnoty štandardizovaných reziduí vypočítať a následne je možné ich rozdelenie testovať. Ponuka testov normality v Minitabe je na obr.1.43, štandardizované reziduá boli zapísané v stĺpci SRES1. Excel v tomto prípade riešenie neposkytuje.



Obr. 1.43 Postup spustenia testov normality v Minitabe

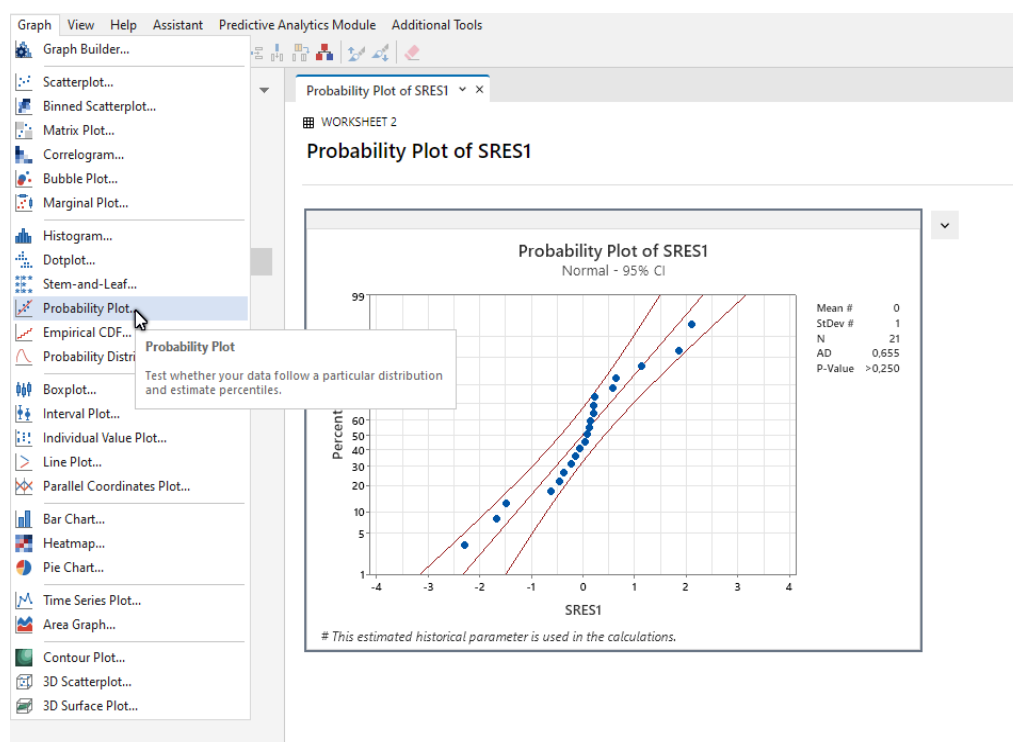
Výstup, t.j. normálny pravdepodobnostný graf spolu s vypočítanou hodnotou testovacej štatistiky, p-hodnotou, parametrami rozdelenia a rozsahom súboru môžeme vidieť na obr. 1.44. Pre analyzovaných 21 hodnôt bola vypočítaná stredná hodnota $5,1 \cdot 10^{-16}$, štandardná odchýlka 1,032. Na základe uvedeného, analyzované dáta (štandardizované reziduá) spĺňajú parametre štandardizovaného normálneho rozdelenia $N(0;1)$. Hodnota testovacej štatistiky je vyjadrená ako korelačný koeficient $r = 0,966$. Pretože vypočítaná p-hodnota $> 0,1$, môžeme prijať

nulovú hypotézu H_0 : Dáta pochádzajú z normálneho rozdelenia a zamietnuť alternatívnu hypotézu H_1 : Dáta nepochádzajú z normálneho rozdelenia.



Obr. 1.44 Normálny pravdepodobnostný graf s výstupom zvoleného Ryan-Joiner testu

K rovnakému záveru sa dostaneme pomocou grafického zobrazenia normálneho pravdepodobnostného grafu, kde je prednastavený Anderson-Darlingov test (obr.1.45).

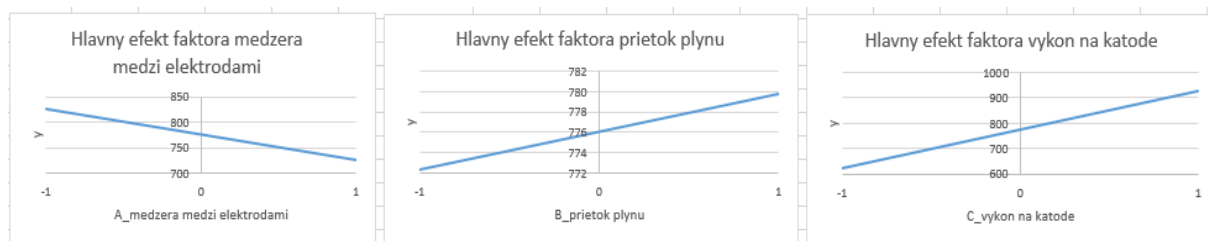


Obr. 1.45 Normálny pravdepodobnostný graf s výstupom Anderson-Darling testu

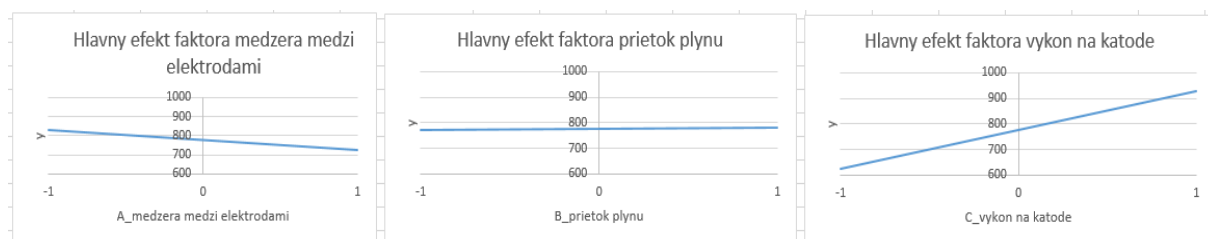
Grafy hlavných efektov faktorov a interakcií faktorov

Cieľom DOE býva porovnať veľkosť vplyvu faktorov a ich interakcií na skúmanú odozvu y . Porovnávame, ktoré faktory a interakcie medzi faktormi majú najvýznamnejší, t.j. najväčší vplyv na výslednú nameranú hodnotu experimentu. Veľkosť vplyvu faktora na odozvu kvantifikujeme efektom faktora. Hlavný efekt faktora A vyjadruje zmenu priemernej hodnoty odozvy, ktorá sa prejaví pri zmene úrovne tohto faktora z dolnej na hornú, keď ostatné faktory zostávajú konštantné, čo zapíšeme vzťahom $ef(A) = \bar{y}_{A+} - \bar{y}_{A-}$. Ak je hlavný efekt faktora kladný, potom pri zmene faktora z dolnej na hornú úroveň odozva y narastie, ak je záporný, hodnota odozvy y klesne. Hlavný efekt faktora zobrazujeme úsečkou.

Grafy hlavných efektov skúmaných faktorov sú na obr. 1.46. Hlavný efekt faktora $A_{\text{medzera medzi elektródami}}$ je záporný, t.j. zmenou z dolnej na hornú úroveň faktora nameraná rýchlosť nitrídového leptania klesla. Hlavné efekty faktorov $B_{\text{prietok plynu}}$ a $C_{\text{výkon na katóde}}$ sú kladné, t.j. zmenou z dolnej na hornú úroveň faktora rýchlosť nitrídového leptania narastla. Veľkosť vplyvu faktora na odozvu je v prípade faktora $B_{\text{prietok plynu}}$ v porovnaní s ostatnými faktormi zanedbateľný. Pre vizuálne porovnanie veľkosti vplyvov faktorov na odozvu je potrebné na y-ovej osi zvoliť rovnaký rozsah, obr. 1.47.

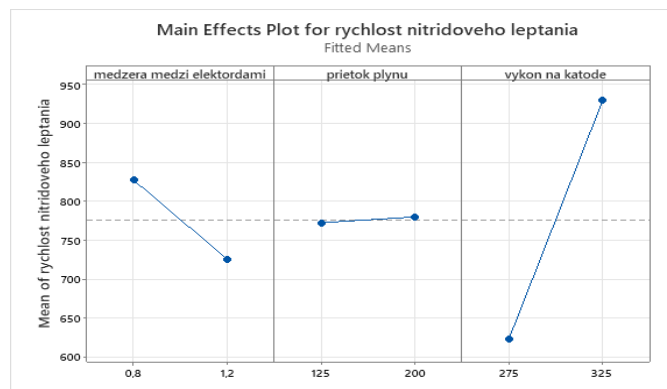


Obr. 1.46 Zobrazenie grafov hlavných efektov skúmaných faktorov v Exceli



Obr. 1.47 Vizuálne porovnanie veľkosti hlavných efektov skúmaných faktorov umožňuje nastavenie rovnakej mierky y-ovej osi v Exceli

Prehľadné porovnanie veľkosti vplyvov skúmaných faktorov na odozvu zobrazuje výstup Minitabu na obr. 1.48. Výpočet priemerných hodnôt odozvy na jednotlivých úrovniach faktorov v Exceli spolu vypočítanými veľkosťami efektov sú na obr. 1.49.



Obr. 1.48 Grafy hlavných efektov skúmaných faktorov v Minitabe

	A	B	C	H	I	J	K	L
1	A_t1	B_t2	C_t3	yi				
2	-1	-1	-1	550			hlavny efekt faktora	
3	-1	-1	-1	604			-1	=AVERAGE(H2:H3;H6:H7;H10:H11;H14:H15)
4	1	-1	-1	669			1	=AVERAGE(H4:H5;H8:H9;H12:H13;H16:H17)
5	1	-1	-1	650			ef(A)=	=L4-L3
6	-1	1	-1	633				
7	-1	1	-1	601				
8	1	1	-1	642			hlavny efekt faktora	
9	1	1	-1	635			-1	=AVERAGE(H2:H5;H10:H13)
10	-1	-1	1	1037			1	=AVERAGE(H6:H9;H14:H17)
11	-1	-1	1	1052			ef(B)=	=L10-L9
12	1	-1	1	749				
13	1	-1	1	868				
14	-1	1	1	1075			hlavny efekt faktora	
15	-1	1	1	1063			-1	=AVERAGE(H2:H9)
16	1	1	1	729			1	=AVERAGE(H10:H17)
17	1	1	1	860			ef(C)=	=L16-L15

	A	B	C	H	I	J	K	L	M
1	A_t1	B_t2	C_t3	yi					
2	-1	-1	-1	550			hlavny efekt faktora A_medzera		
3	-1	-1	-1	604			-1	826,875	
4	1	-1	-1	669			1	725,25	
5	1	-1	-1	650			ef(A)=	-101,625	
6	-1	1	-1	633					
7	-1	1	-1	601					
8	1	1	-1	642			hlavny efekt faktora B_prietok		
9	1	1	-1	635			-1	772,375	
10	-1	-1	1	1037			1	779,75	
11	-1	-1	1	1052			ef(B)=	7,375	
12	1	-1	1	749					
13	1	-1	1	868					
14	-1	1	1	1075			hlavny efekt faktora C_vykon na katode		
15	-1	1	1	1063			-1	623	
16	1	1	1	729			1	929,125	
17	1	1	1	860			ef(C)=	306,125	

Obr. 1.49 Výpočet hlavných efektov faktorov v Exceli. Hlavný efekt faktora je rozdiel v priemernej odozve, keď sa faktor mení z dolnej úrovne na hornú, pričom ostatné faktory zostávajú konštantné.

Okrem hlavného efektu faktorov skúmame veľkosť vplyvu interakcie faktorov na odozvu, kedy nehodnotíme len samotný vplyv faktora, ale aj to, ako jeho účinok ovplyvňuje kombinácia s inými faktormi. Grafy interakcie faktorov v dvojúrovňových experimentoch zobrazujú pomocou dvoch úsečiek pre faktor a jeho úroveň (na vodorovnej osi) priemernú nameranú hodnotu odozvy pre obe úrovne druhého faktora (v legende). Ak sú úsečky v grafe paralelné, potom neexistuje interakcia medzi faktormi. Vplyv jedného faktora na odozvu je rovnaký pre všetky úrovne druhého faktora. Rôznobežné úsečky naznačujú, že existuje interakcia medzi faktormi, t.j. vplyv jedného faktora závisí od úrovne druhého faktora. Čím väčší je rozdiel v sklone úsečiek, tým silnejšia je interakcia. Výpočet priemerov odozvy pre zvolené úrovne faktorov je v nasledujúcich tabuľkách na obr. 1.50 - 1.52.

	A	B	C	D	E	F
1	A_t1	B_t2	C_t3	yi		
2	-1	-1	-1	550		
3	-1	-1	-1	604		
4	1	-1	-1	669		
5	1	-1	-1	650		
6	-1	1	-1	633		
7	-1	1	-1	601		
8	1	1	-1	642		
9	1	1	-1	635		
10	-1	-1	1	1037		
11	-1	-1	1	1052		
12	1	-1	1	749		
13	1	-1	1	868		
14	-1	1	1	1075		
15	-1	1	1	1063		
16	1	1	1	729		
17	1	1	1	860		
18						
19	efekt interakcie faktora A a C - pocitame efekt faktora C pri konstantnej urovni faktora A alebo naopak					
20						
21				A		
22				0,8	1,2	
23				-1	1	
24	C	275	-1	=AVERAGE(D2:D3;D6:D7)	=AVERAGE(D4:D5;D8:D9)	=AVERAGE(E25;D24)
25		325	1	=AVERAGE(D10;D11;D14;D15)	=AVERAGE(D12;D13;D16;D17)	=AVERAGE(D25;E24)
26				ef(AC)= =F24-F25		

19	efekt interakcie faktora A a C - pocitame efekt faktora C pri konstantnej urovni faktora A alebo naopak					
20						
21				A		
22				0,8	1,2	
23				-1	1	
24	C	275	-1	597	649	699,25
25		325	1	1056,75	801,5	852,875
26				ef(AC)=		-153,625

Obr. 1.50 Tabuľka s výpočtom priemerov odoziev pre zvolené úrovne faktorov A a C

28	efekt interakcie faktora A a B - pocitame efekt faktora B pri konstantnej urovni faktora A alebo naopak					
29						
30				A		
31				0,8	1,2	
32				-1	1	
33	B	125	-1	=AVERAGE(D2;D3;D10;D11)	=AVERAGE(D4;D5;D12;D13)	=AVERAGE(E34;D33)
34		200	1	=AVERAGE(D6;D7;D14;D15)	=AVERAGE(D8;D9;D16;D17)	=AVERAGE(D34;E33)
35						ef(AB)= =F33-F34

28	efekt interakcie faktora A a B - pocitame efekt faktora B pri konstantnej urovni faktora A alebo naopak					
29						
30				A		
31				0,8	1,2	
32				-1	1	
33	B	125	-1	810,75	734	763,625
34		200	1	843	716,5	788,5
35						ef(AB)= -24,875

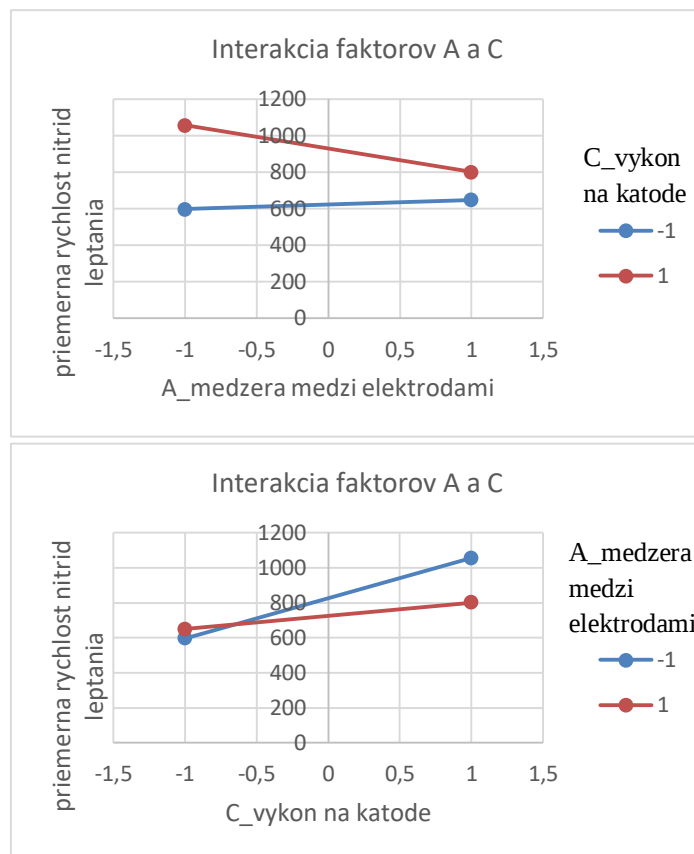
Obr. 1.51 Tabuľka s výpočtom priemerov odoziev pre zvolené úrovne faktorov A a B

37	efekt interakcie faktora B a C - pocitame efekt faktora C pri konstantnej urovni faktora B alebo naopak					
38						
39				B		
40				125	200	
41				-1	1	
42	C	275	-1	=AVERAGE(D2;D3;D4;D5)	=AVERAGE(D6;D7;D8;D9)	=AVERAGE(E43;D42)
43		325	1	=AVERAGE(D10;D11;D12;D13)	=AVERAGE(D14;D15;D16;D17)	=AVERAGE(D43;E42)
44						ef(BC)= =F42-F43

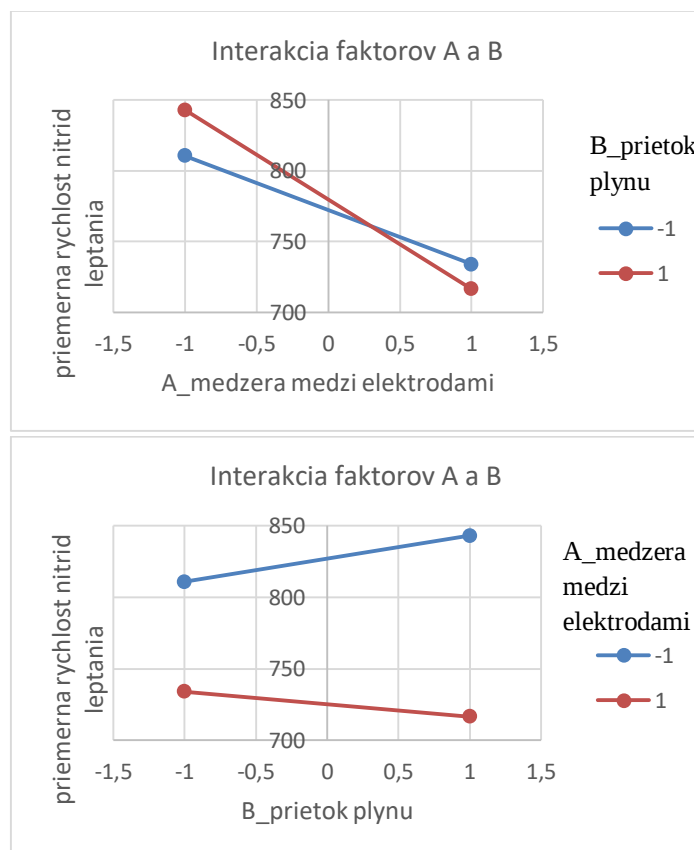
37	efekt interakcie faktora B a C - pocitame efekt faktora C pri konstantnej urovni faktora B alebo naopak					
38						
39				B		
40				125	200	
41				-1	1	
42	C	275	-1	618,25	627,75	775
43		325	1	926,5	931,75	777,125
44						ef(BC)= -2,125

Obr. 1.52 Tabuľka s výpočtom priemerov odoziev pre zvolené úrovne faktorov B a C

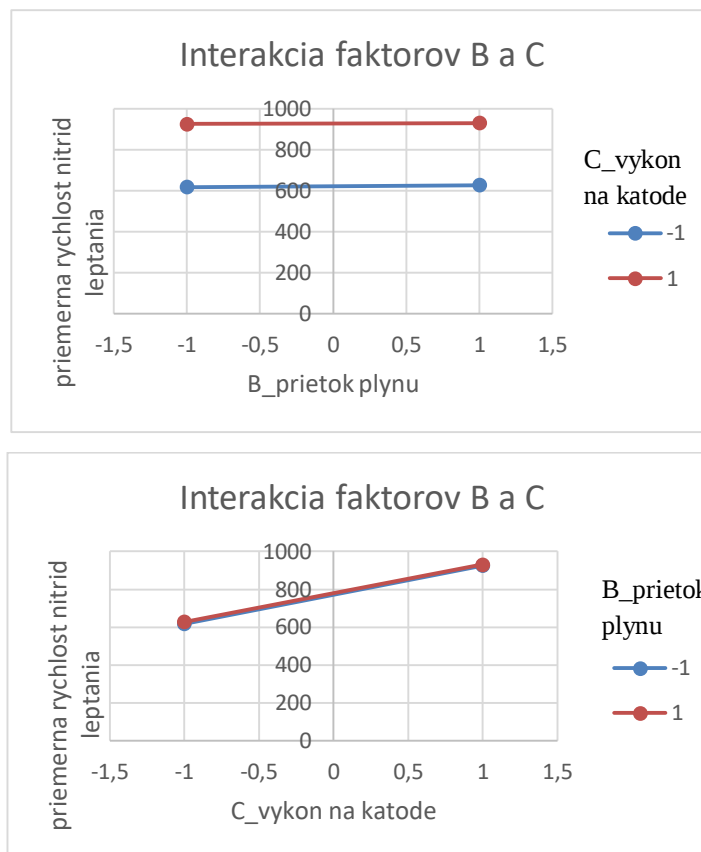
Pri zobrazovaní grafov interakcie faktorov po označení príslušnej celej tabuľky s vypočítanými hodnotami priemerov odoziev pre zvolené úrovne faktorov volíme ponuku *Vložiť* → *Závislosť*: *bodový graf s priamkami a značkami*. Po naformátovaní grafu (napr. v ponuke *Zdrojové údaje*) vieme graf interakcií jednoduchým spôsobom skopírovať a zmeniť prepnutím riadku a stĺpca v ponuke *Nástroje pre grafy* v časti *Návrh*, obr. 1.53. Aj v prípade interakcie dvoch faktorov každá úsečka zobrazuje efekt faktora, ktorého úrovne sú vynesené na vodorovnej osi, pričom legenda poskytuje informáciu o tom, na akej úrovni bol nastavený druhý faktor vystupujúci v uvažovanej interakcii.



Obr. 1.53 Grafy interakcie faktorov *A_medzera medzi elektródami* a *C_výkon na katóde*

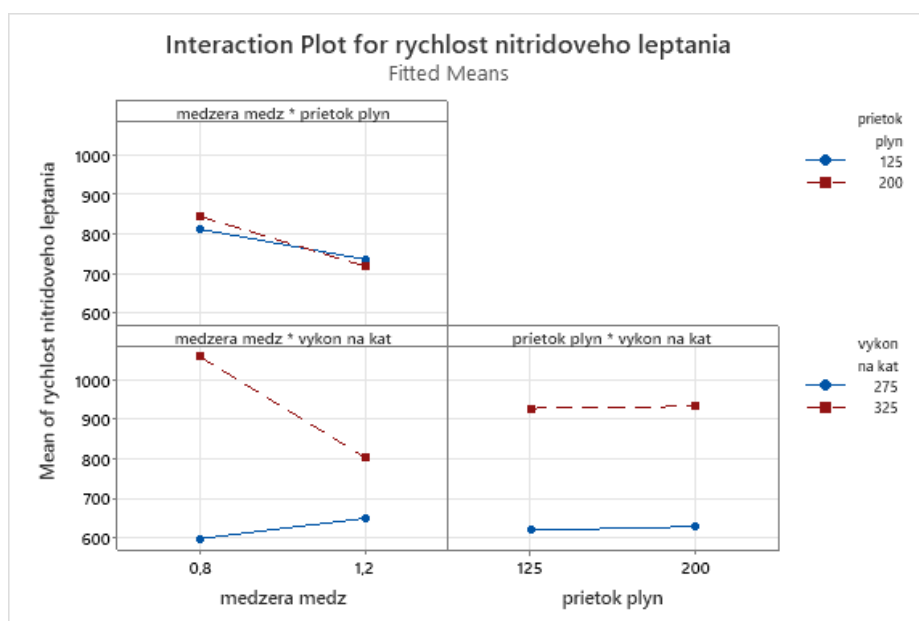


Obr. 1.54 Grafy interakcie faktorov *A_medzera medzi elektródami* a *B_prietok plynu*



Obr. 1.55 Grafy interakcie faktorov $B_{\text{prietok plynu}}$ a $C_{\text{výkon na katóde}}$

Prehľadný graf s interakciami všetkých kombinácií dvojíc faktorov, v ktorom je možné jednoducho porovnávať rôznobežnosť zobrazovaných úsečiek, t.j. veľkosť interakcie uvažovaných faktorov, poskytuje Minitab na obr. 1.56.



Obr. 1.56 Grafy interakcií faktorov v Minitabe

Z grafu interakcie faktorov vyplýva:

- 1) medzi B_prietokom plynu a C_výkonom na katóde nie je prítomná interakcia – zobrazené úsečky sú rovnobežné;
- 2) medzi A_medzerou medzi elektródami a C_výkonom na katóde je pozorovaná najväčšia interakcia – uhol medzi úsečkami je najväčší;
- 3) najväčšiu odozvu možno očakávať pri nastavení faktora A_medzera medzi elektródami na dolnú úroveň, faktora C_výkon na katóde na hornú roveň a faktor B_prietok plynu na hornú úroveň, pričom vplyv faktora B možno považovať za zanedbateľný.
- 4) najmenšiu odozvu možno očakávať pri nastavení faktora A_medzera medzi elektródami na dolnú úroveň, faktora C_výkon na katóde na dolnú roveň, vplyv faktora B_prietok plynu možno zanedbať.

Pareto graf – porovnanie veľkosti efektov faktorov a interakcií

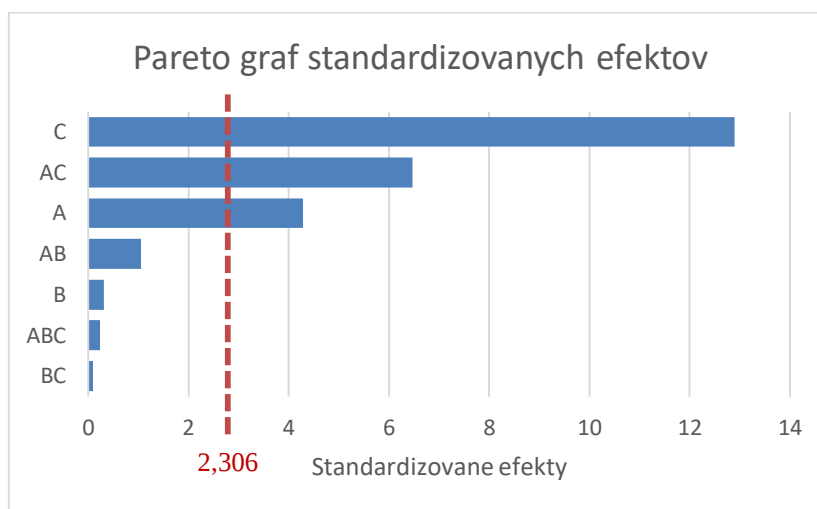
Na vizuálne porovnanie veľkosti vplyvu jednotlivých faktorov a ich interakcií na výslednú premennú (odozvu) experimentu je vhodné použiť Pareto graf, vid' obr.1.57 alebo obr.1.58. Pareto graf zobrazuje absolútne hodnoty štandardizovaných veľkostí efektov zoradených zostupne ako pruhový graf. Štandardizované efekty sú t -štatistiky, ktoré testujú nulovú hypotézu, že regresný koeficient je 0. Zobrazované pruhy (alebo stĺpce) sú zoradené podľa veľkosti, preto je ľahké identifikovať faktory a interakcie s najvýznamnejším vplyvom. Platí pravidlo, čím väčší je pruh (alebo stĺpec) pre určitý faktor alebo interakciu, tým väčší je jeho vplyv na odozvu. Typicky býva v grafe zobrazená aj kritická hodnota CV ako referenčná čiara, ktorá odlišuje významné efekty od nevýznamných. Pri výpočte hodnoty CV rozlišujeme, či vieme určiť stupne voľnosti pre experimentálnu chybu, t. j. rozlišujeme či experiment mal alebo nemal replikácie.

Kritická hodnota pre experiment bez replikácií sa počíta podľa vzťahu:

$CV = PSE \cdot t_{m/3}^{kv}(1 - \alpha)$, kde PSE je pseudoštandardná chyba (Lenth's PSE) založená na skutočnosti, že smerodajnú odchýlku vzorky z rozdelenia $N(0, \sigma^2)$ možno odhadnúť ako 1,5 násobok mediánu absolútnych hodnôt. Významné efekty sú pri spresňujúcich výpočtoch vylúčené ako extrémne hodnoty prevyšujúce 2,5 násobok mediánu.

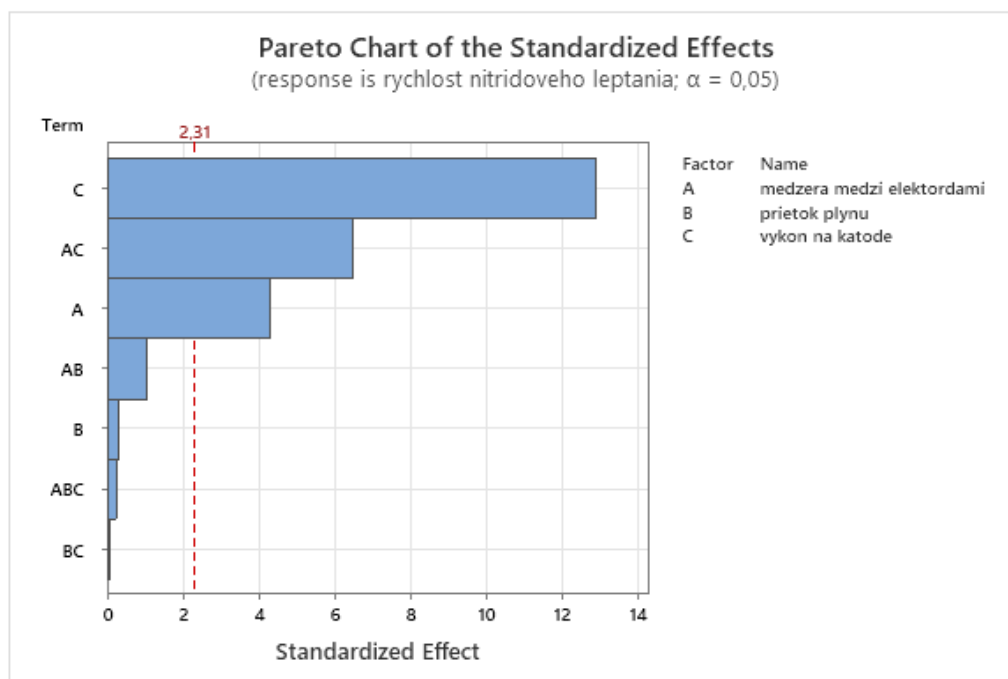
Kvantil Studentovho t -rozdelenia $t_{m/3}^{kv}(1 - \alpha)$ má počet stupňov voľnosti $df = \frac{m}{3}$, pričom m je celkový počet odhadov efektov a interakcií.

Pre experiment s replikáciami je kritická hodnota CV rovná kvantilu Studentovho t -rozdelenia $t_{(m-1).2^k}^{kv}(1 - \alpha/2)$, kde k je počet faktorov a m je počet replikácií v každom experimentálnom bode. Pre zvolené $\alpha = 0,05$ je 97,5% kvantil $t_{(2-1).2^3}^{kv}(1 - \alpha/2) = T.INV(0,975;8)=2,306004$.



Obr. 1.57 Pareto graf štandardizovaných efektov v Exceli

Poznámka: Z Pareto grafu, ktorý zobrazuje absolútne hodnotu štandardizovaných efektov, je možné určiť, ktoré efekty sú veľké, ale nemôžeme určiť, ktoré efekty zvyšujú alebo znižujú odozvu. Veľkosť a smer účinkov je možné zistiť z tabuľky, ktorá obsahuje bodové odhady koeficientov. Ak sú niektoré faktory alebo interakcie významné, je vhodné sa pri ďalšom skúmaní alebo hľadaní optimálneho riešenia na ne zamerať a ostatné nevýznamné členy zanedbať.



Obr. 1.58 Pareto graf štandardizovaných efektov v Minitabe

Zhrnutie:

Z grafov hlavných efektov vyplýva, že v prípade, ak by nebola prítomná interakcia faktorov, potom by na rýchlosť nitridového leptania najviac vplýval faktor C_výkon na katóde, pričom tento efekt faktora je pozitívny, t.j. faktor zvyšuje odozvu. Rovnako pozitívny vplyv má aj faktor B_prietok plynu, ale veľkosť tohto vplyvu na odozvu je v porovnaní s ostatnými faktormi zanedbateľný. Naopak vplyv faktora A_medzera medzi elektródami na odozvu je negatívny, pretože faktor znižuje odozvu. Vplyv tohto faktora na rýchlosť nitridového leptania je menší ako vplyv faktora C_výkon na katóde a súčasne väčší ako vplyv faktora B_prietok plynu.

Z grafov interakcií faktorov však vyplýva, že možno očakávať vzájomné pôsobenie faktorov A_medzera medzi elektródami a B_prietok plynu a rovnako faktorov A_medzera medzi elektródami a C_výkon na katóde na odozvu, preto analýzu vplyvu faktorov na odozvu nemožno považovať za smerodajnú. Interakcia faktorov B_prietok plynu a C_výkon na katóde nebude prítomná, pretože zobrazené úsečky sú rovnobežné.

Z grafov interakcie faktorov A_medzera medzi elektródami a C_výkon na katóde na obr.1.53 vyplýva, že odozva bola vždy vyššia pri hornej úrovni faktora C_výkon na katóde, t.j. pri zvolenom vyššom výkone na katóde bez ohľadu na zvolenú úroveň faktora A_medzera medzi elektródami. Zvýšenie C_výkonu na katóde má oveľa väčší vplyv na rýchlosť leptania, keď je A_medzera medzi elektródami menšia. Ak nastavíme C_výkon na katóde na dolnú úroveň, potom zmena A_medzery medzi elektródami vplýva na zmenu hodnoty odozvy minimálne. V hornom grafe modrá úsečka zobrazuje efekt faktora A_medzera medzi elektródami, keď je C_výkon na katóde nastavený na dolnú úroveň -1. Zvýšenie A_medzery medzi elektródami vedie len k miernemu zvýšeniu priemernej rýchlosti nitridového leptania. Tento nárast je kladný, ale nie je výrazný. Červená úsečka zobrazuje efekt faktora A_medzera medzi elektródami, keď je C_výkon na katóde nastavený na hornú úroveň +1. Zvýšenie A_medzery medzi elektródami vedie k výraznejšiemu poklesu rýchlosti nitridového leptania. Tento efekt je silnejší ako pri zvolenej nižšej hodnote výkonu na katóde. V dolnom grafe modrá úsečka zobrazuje efekt faktora C_výkon na katóde, keď je faktor A_medzera medzi elektródami nastavený na dolnú úroveň -1. Červená úsečka zobrazuje efekt faktora C_výkon na katóde, keď je faktor A_medzera medzi elektródami nastavený na hornú úroveň +1. Keď je A_medzera medzi elektródami nastavená na dolnú úroveň -1, zvýšenie výkonu na katóde vedie

k väčšiemu nárastu rýchlosti nitridového leptania ako pri nastavení A_medzery medzi elektródami na hornú úroveň +1.

Najväčšiu hodnotu odozvy sme podľa grafu interakcie faktorov A a C dosiahli, ak bola zvolená medzera medzi elektródami minimálna a výkon na katóde maximálny.

Z grafov interakcie faktorov A_medzera medzi elektródami a B_prietok plynu na obr.1.54 vyplýva, že odozva bola vždy vyššia pri dolnej úrovni faktora A_medzera medzi elektródami, t.j. pri zvolenom nižšej medzere medzi elektródami bez ohľadu na zvolenú úroveň faktora B_prietok plynu. Zväčšenie medzery medzi elektródami vždy znižuje rýchlosť nitridového leptania, v prípade vyššieho prietoku plynu je táto zmena výraznejšia. Ak nastavíme medzeru medzi elektródami na dolnú úroveň, potom zvýšenie prietoku plynu zvyšuje rýchlosť nitridového leptania, ak nastavíme medzeru medzi elektródami na hornú úroveň, potom zvýšenie prietoku plynu zníži rýchlosť nitridového leptania, pričom táto zmena je menšia. V hornom grafe modrá úsečka zobrazuje efekt faktora A_medzera medzi elektródami, keď je B_prietok plynu nastavený na dolnú úroveň -1. Červená úsečka zobrazuje efekt faktora A_medzera medzi elektródami, keď je B_prietok plynu nastavený na hornú úroveň +1.

V dolnom grafe modrá úsečka zobrazuje efekt faktora B_prietok plynu, keď je faktor A_medzera medzi elektródami nastavený na dolnú úroveň -1. Červená úsečka zobrazuje efekt faktora B_prietok plynu, keď je faktor A_medzera medzi elektródami nastavený na hornú úroveň +1. Nastavenie úrovne faktora A_medzera medzi elektródami určuje, či bude efekt faktora pozitívny (pre úroveň -1) alebo negatívny (pre úroveň +1).

Najväčšiu hodnotu odozvy sme podľa grafu interakcie faktorov A a B dosiahli, ak bola zvolená medzera medzi elektródami minimálna a prietok plynu maximálny.

Z grafov interakcie faktorov B_prietok plynu a C_výkon na katóde na obr.1.55 je vidieť, že interakcia medzi týmito faktormi nie je prítomná. Na odozvu vplýva iba faktor C_výkon na katóde. Jeho efekt je kladný.

Overenie regresného modelu pomocou kontrolného merania

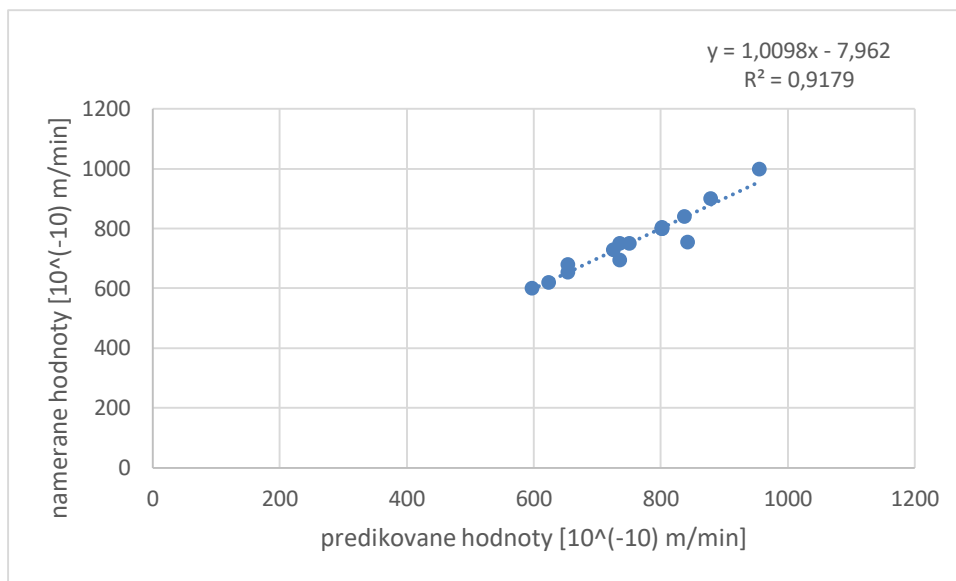
Po nájdení regresného modelu je potrebné vykonať kontrolné meranie, pričom volíme body z daného experimentálneho priestoru, kde preferujeme úrovne faktorov, ktoré neboli súčasťou pôvodných kombinácií úrovní faktorov v pôvodných experimentálnych meraniach. Ak toto nie je možné realizovať, náhodne vyberieme rovnaké body experimentálneho priestoru, ktoré boli použité pri hľadaní regresného modelu. Týmto spôsobom získame nové hodnoty pre hodnoty odozvy y .

Po vykonaní kontrolných meraní pre vybrané úrovne faktorov boli získané nasledovné hodnoty rýchlosti nitridového leptania [10^{-10} m/min]:

$f(0,8;125;275)=602$; $f(1,2;200;325)=804$; $f(0,9;150;300)=800$; $f(1;130;280)=681$;
 $f(0,9;190;310)=900$; $f(1,1;130;320)=755$; $f(0,8;140;290)=750$; $f(1;190;310)=840$;
 $f(1;160;275)=621$; $f(1,1;180;300)=750$; $f(1,2;160;300)=730$; $f(0,9;130;300)=800$;
 $f(1;180;280)=654$; $f(0,8;190;290)=695$; $f(0,9;150;320)=1000$;

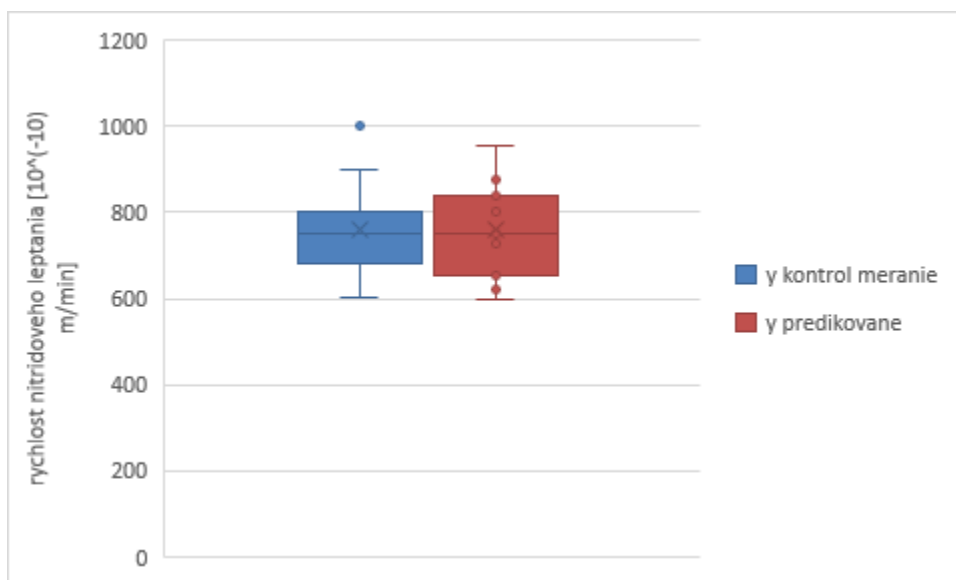
Kontrolne meranie:					predikcia_povodne jednotky				
p.c.	x1	x2	x3	Y		x1	x2	x3	y predik
1	0,8	125	275	602		0,8	125	275	597
2	1,2	200	325	804		1,2	200	325	801,5
3	0,9	150	300	800		0,9	150	300	801,46875
4	1	130	280	681		1	130	280	653,6125
5	0,9	190	310	900		0,9	190	310	878,05625
6	1,1	130	320	755		1,1	130	320	842,38125
7	0,8	140	290	750		0,8	140	290	734,925
8	1	190	310	840		1	190	310	837,2875
9	1	160	275	621		1	160	275	623
10	1,1	180	300	750		1,1	180	300	750,65625
11	1,2	160	300	730		1,2	160	300	725,25
12	0,9	130	300	800		0,9	130	300	801,46875
13	1	180	280	654		1	180	280	653,6125
14	0,8	190	290	695		0,8	190	290	734,925
15	0,9	150	320	1000		0,9	150	320	954,64375

Porovnaním hodnôt získaných z regresného modelu (predikovaných hodnôt) s kontrolným meraním overujeme, či je regresný model dostatočne presný na predikciu hodnôt nitridového leptania. Vizuálnou kontrolou v bodovom grafe na obr.1.59 možno pozorovať, že modelované hodnoty zodpovedajú skutočným kontrolným nameraným hodnotám. Okrem toho koeficient determinácie $R^2 = 91,79\%$ ukazuje, že lineárny model veľmi dobre fituje dané údaje. Pearsonov koeficient korelácie $r = \sqrt{R^2} = 0,958055$ ako miera tesnosti lineárnej závislosti medzi modelovanými a kontrolnými nameranými hodnotami má hodnotu blízku 1, čo naznačuje vysoký stupeň priamej lineárnej závislosti medzi nimi.



Obr. 1.59 Validácia regresného modelu

Na obr. 1.60 sú zobrazené Box ploty, ktoré porovnávajú rozdelenie hodnôt kontrolného merania s predikovanými hodnotami, ktoré zodpovedajú rovnakým úrovňam vstupných parametrov. V kontrolnom meraní sme zaznamenali jednu extrémnu hodnotu $1000 \cdot 10^{-10}$ m/min. Vzhľadom na namerané rýchlosti nitridového leptania pri realizácii experimentu a bodový graf na obr. 1.59 nie je potrebné túto hodnotu ďalej riešiť.



Obr. 1.60 Box ploty porovnávajúce hodnoty kontrolného

Okrem bežného a jednoduchého vizuálneho prístupu pomocou bodového grafu a porovnania Box plotov a štatistík r a R^2 uvedených vyššie, existuje aj kontrola pomocou tzv. porovnania parametrov smernice a priesečníka y -ovej osi po preložení vynesných bodov

priamkou. Ak majú byť modelované hodnoty v porovnaní s kontrolnými nameranými v pomere 1:1, fitovaná priamka by mala sklon 45° a pretínať os y v nule. Vo výslednom lineárnom regresnom modeli $y = b_0 + b_1x$ preto očakávame, že parameter $b_1 = 1$ a $b_0 = 0$. Na základe tejto myšlienky testujeme, či regresný koeficient b_1 je rovný jednej a lokujúca konštanta b_0 je rovná nule.

Testujeme hypotézy $H_0: b_1 = 1 \leq H_1: b_1 \neq 1$. Nech hladina významnosti $\alpha = 0,05$. Z výstupu regresnej analýzy v Exceli na obr. 1.61 môžeme použiť bodový odhad regresného koeficientu b_1 a jeho 95% interval spoľahlivosti.

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0,9580551					
R Square	0,9178696					
Adjusted R Square	0,9115519					
Standard Error	31,376231					
Observations	15					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	143028,3173	143028,3	145,2849	1,9813E-08	
Residual	13	12798,08272	984,4679			
Total	14	155826,4				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-7,962006	64,12743812	-0,12416	0,903089	-146,500913	130,576902
X Variable 1	1,009802	0,083777228	12,05342	1,98E-08	0,828812292	1,19079169

Obr. 1.61 Výstup regresnej analýzy pri lineárnom regresnom modeli v Exceli

Testovacia štatistika $t(b_1) = \frac{\hat{b}_1 - 1}{SE((b_1))} \sim t(n - 2)$. Vzhľadom na obojstrannú alternatívu p hodnotu učíme podľa vzťahu $p(b_1) = 2 * \min\{F(t_0(b_1)); 1 - F(t_0(b_1))\}$. Vzhľadom na zvolené riziko intervalového odhadu α , obojstranný $(1 - \alpha) \cdot 100\%$ interval spoľahlivosti určíme podľa vzťahu $\hat{b}_1 \pm \Delta$, kde $\Delta = SE(\hat{b}_1) \cdot t_{n-2}^{kv} \left(1 - \frac{\alpha}{2}\right)$.

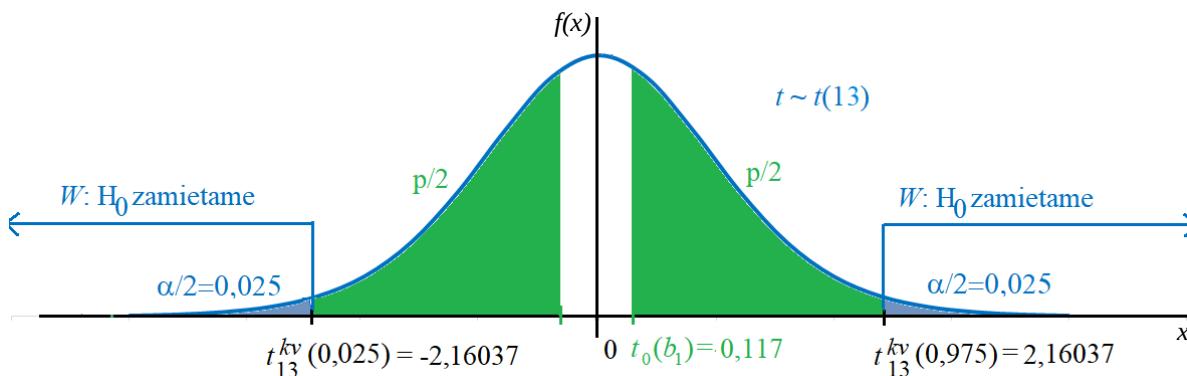
Výpočet uvedených hodnôt v Exceli je zobrazený na nasledujúcom obr. 1.62. Nie je možné použiť údaje z riadku pre bodový odhad koeficientu $\hat{b}_1$ s výnimkou intervalu spoľahlivosti, pretože testovacia štatistika i p hodnota testujú nulovosť regresného koeficientu b_1 , t.j. hypotézu $H_0: b_1 = 0$ oproti hypotéze $H_1: b_1 \neq 0$.

	C	D	E
57		<i>Coefficients</i>	<i>Standard Error</i>
58	Intercept	-7,96200563754076	64,1274381168403
59	X Variable 1	1,00980199012169	0,0837772283084262
60			
61	H0: b1=1 ><H1: b1 ≠ 1		
62	t0= =(D59-1)/E59		
63	t ₁₃ ^{kv} (0,975)= =T.INV(0,975;13)		
64	p= =2*MIN(T.DIST(D62;13;1);1-T.DIST(D62;13;1))		
65	delta= =E59*D63		
66	95%IS pre b1:		
67	=D59-D65	=D59+D65	

61	H0: b1=1 ><H1: b1 ≠ 1	
62	t0=	0,1170006
63	t ₁₃ ^{kv} (0,975)=	2,1603687
64	p=	0,9086478
65	delta=	0,1809897
66	95%IS pre b1:	
67	0,828812292	1,1907917

Obr. 1.62 Testovanie regresného koeficienta b_1 v Exceli

Grafické riešenie je na obr. 1.63



Obr. 1.63 Grafické riešenie pre testovanie hypotézy o koeficiente b_1

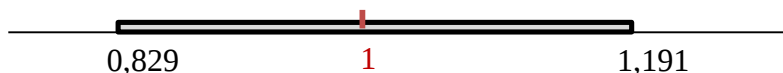
Kritická oblasť $W = (-\infty; -2,16) \cup (2,16, \infty)$, kde 97,5% kvantil $t_{13}^{kv}(0,975) = T.INV(0,975; 13) = 2,16037$. Platí $|t_0(b_1)| = \frac{1,009802-1}{0,083777228} = 0,1170006 < 2,16037 = t_{13}^{kv}(0,975)$.

Z grafického riešenia pre test hypotézy o koeficiente $H_0 : b_1 = 1$ na obr. 1.63 realizácia testovacej štatistiky $t_0(b_1) = 0,117 \notin W$, preto testovanú hypotézu $H_0 : b_1 = 1$ nezamietame. Naopak hypotézu $H_1 : b_1 \neq 1$ zamietame.

P-hodnota vypočítaná podľa vzťahu $p(b_1) = 2 * \min\{F(t_0(b_1)); 1 - F(t_0(b_1))\}$ je rovná $p(b_1) = 2 * \min(T. \text{DIST}(0,117; 13; 1); 1 - T. \text{DIST}(0,117; 13; 1)) = 0,90865$
 $p = 0,90865 > 0,05 = \alpha$, preto $H_0 : b_1 = 1$ nezamietame.

Graficky p-hodnota predstavuje dvojnásobok veľkosti plochy pod krivkou hustoty rozdelenia Studentovho rozdelenia $t(13)$ od realizácie testovacej štatistiky $t_0(b_3)$ vpravo.

95% IS pre koeficient b_1 je podľa vzťahu $\hat{b}_1 \pm \Delta$ rovný $1,009802 \pm 0,1509897$. Rovnako z výstupu regresnej analýzy na obr. 1.61 vidíme, že 95% IS pre koeficient b_1 je rovný $(0,829; 1,191)$



Obr. 1.64 1 patrí do 95% IS pre koeficient b_1

Keďže $1 \in (0,829; 1,191)$, so spoľahlivosťou 95% môžeme hypotézu $H_0 : b_1 = 1$ prijať.

Záver:

Na hladine významnosti $\alpha = 0,05$ možno prijať hypotézu, že regresný koeficient b_1 je rovný jednej.

V nasledujúcom kroku testujeme hypotézy $H_0 : b_0 = 0 \leq H_1 : b_0 \neq 0$ pri zvolenej hladine významnosti $\alpha = 0,05$. Môžeme použiť výstup regresnej analýzy na obr. 1.61, konkrétne riadok prislúchajúci lokujúcej konštante b_0 označený „Intercept“, vid' obr. 1.65.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-7,96200563754076	64,1274381168403	-0,124159109	0,903088541	-146,50091296442	130,576901689338
X Variable 1	1,00980199012169	0,0837772283084262	12,053418458	1,981304394	0,828812291958836	1,19079168828454

Obr. 1.65 Tabuľka s bodovými odhadmi koeficientov a im prislúchajúcimi hodnotami

Testovacia štatistika $t(b_0) = \frac{\hat{b}_0}{SE((b_0))} \sim t(n - 2)$. Vzhľadom na obojstrannú alternatívu p hodnotu učíme podľa vzťahu $p(b_0) = 2 * \min\{F(t_0(b_0)); 1 - F(t_0(b_0))\}$. Vzhľadom na zvolené riziko intervalového odhadu α , obojstranný $(1 - \alpha) \cdot 100\%$ interval spoľahlivosti určíme podľa vzťahu $\hat{b}_0 \pm \Delta$, kde $\Delta = SE(\hat{b}_0) \cdot t_{n-2}^{kv} \left(1 - \frac{\alpha}{2}\right)$.

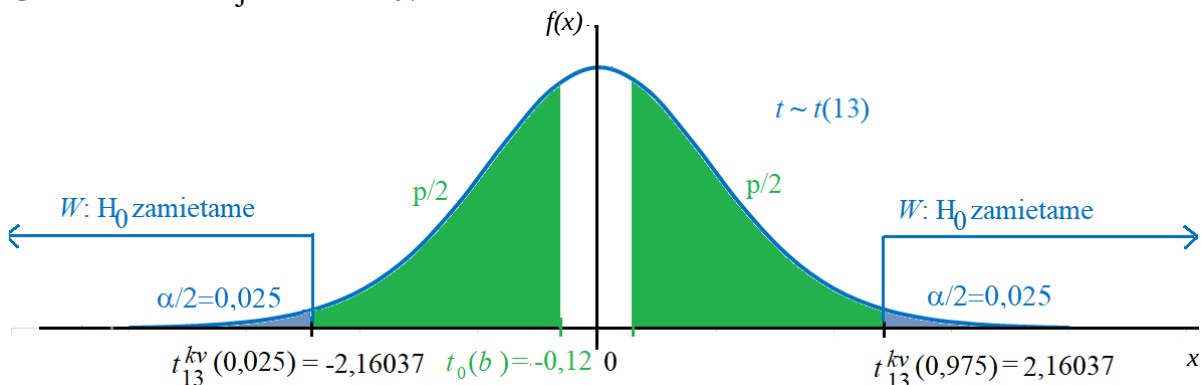
Výpočet uvedených hodnôt v Exceli je zobrazený na nasledujúcom obr. 1.65. Pri testovaní hypotéz možné použiť všetky údaje z riadku pre bodový odhad koeficientu $\hat{b}_0$.

	C	D	E	F	G	H	I
57		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
58	Intercept	-7,96200563754076	64,1274381168403	-0,124159109	0,903088541	-146,50091296442	130,576901689338
59	X Variable 1	1,00980199012169	0,0837772283084262	12,053418458	1,981304394	0,828812291958836	1,19079168828454
60							
61	H0: b0=0 ><H1: b0≠0						
62	t0= =D58/E58						
63	t ₁₃ ^{kv} (0,975)= =T.INV(0,975;13)						
64	p= =2*MIN(T.DIST(D62;13;1);1-T.DIST(D62;13;1))						
65	delta= =E58*D63						
66	95%IS pre b1:						
67	=D58-D65 =D58+D65						

61	H0: b0=0 ><H1: b0≠0	
62	t0= -0,124159	
63	t ₁₃ ^{kv} (0,975)= 2,1603687	
64	p= 0,9030885	
65	delta= 138,53891	
66	95%IS pre b1:	
67	-146,500913	130,5769

Obr. 1.66 Testovanie lokujúcej konštanty b_0 v Exceli

Grafické riešenie je na obr. 1.67



Obr. 1.67 Grafické riešenie pre testovanie hypotézy o koeficiente b_0

Kritická oblasť bude rovnaká, t. j. $W = (-\infty; -2,16) \cup (2,16; \infty)$, kde 97,5% kvantil $t_{13}^{kv}(0,975) = T.INV(0,975; 13) = 2,16037$. Platí $|t_0(b_1)| = \left| \frac{-7,962006}{64,127438} \right| = 0,12416 < 2,16037 = t_{13}^{kv}(0,975)$.

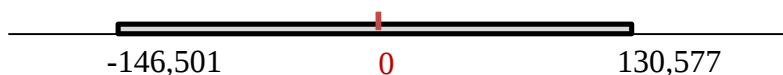
x

Z grafického riešenia pre test hypotézy o koeficiente $H_0 : b_0 = 0$ na obr. 1.67 realizácia testovacej štatistiky $t_0(b_0) = -0,124 \notin W$, preto testovanú hypotézu $H_0 : b_0 = 0$ nezamietame. Naopak hypotézu $H_1 : b_0 \neq 0$ zamietame.

P-hodnota vypočítaná podľa vzťahu $p(b_0) = 2 * \min\{F(t_0(b_0)); 1 - F(t_0(b_0))\}$ je rovná $p(b_0) = 2 * \text{MIN}(T.DIST(-0,124; 13; 1); 1 - T.DIST(-0,124; 13; 1)) = 0,90309$

Vypočítaná i tabuľková hodnota $p = 0,90309 > 0,05 = \alpha$, preto $H_0 : b_0 = 0$ nezamietame. Graficky p-hodnota predstavuje dvojnásobok veľkosti plochy pod krivkou hustoty rozdelenia Studentovho rozdelenia $t(13)$ od realizácie testovacej štatistiky $t_0(b_3)$ vľavo.

95% IS pre koeficient b_0 je podľa vzťahu $\hat{b}_0 \pm \Delta$ rovný $-7,962006 \pm 138,53891$. Vypočítaný 95% IS pre koeficient b_0 i interval uvedený v tabuľke je rovný $(-146,501; 130,577)$.



Obr. 1.68 Číslo 0 patrí do 95% IS pre koeficient b_0

Keďže $0 \in (-146,501; 130,577)$, so spoľahlivosťou 95% môžeme hypotézu $H_0 : b_0 = 0$ prijať.

Záver:

Na hladine významnosti $\alpha = 0,05$ možno prijať hypotézu, že koeficient b_0 je rovný nule. Vzhľadom na to, že je možné prijať hypotézu, že lokujúca konštanta b_0 je rovná nule, to znamená preložená priamka prechádza začiatkom súradnicovej sústavy a súčasne regresný koeficient b_1 , t.j. smernica preloženej priamky je rovná jednej, môžeme prijať tvrdenie, že predikované hodnoty nitridového leptania získané z nájdeného regresného modelu a hodnoty namerané pri kontrolnom meraní pre rovnaké úrovne nami skúmaných parametrov sú v pomere 1:1. Toto možno tvrdiť so spoľahlivosťou 95%.

Predikcia

Na základe predchádzajúcej časti môže byť nájdený regresný model použitý na predikciu hodnôt odozvy alebo na hľadanie optimálneho riešenia. Predikciu i optimalizáciu by sme však mali robiť iba v rámci skúmaného experimentálneho priestoru.

Použitie Excelu na výpočet predikovaných hodnôt je vidieť na nasledujúcom obr. 1.69. Napríklad, ak chceme pomocou nami získaného regresného modelu odhadnúť hodnotu nitridového leptania ak zvolíme faktor A_medzera medzi elektródami na úroveň 0,9 cm, faktor B_prietok plynu na úroveň 150 cm³/min a faktor C_výkon na katóde na úroveň 280W, potom po dosadení do regresného modelu vyjadreného v pôvodných jednotkách, dostaneme hodnotu odozvy 648,2938. 10⁻¹⁰ m/min. Ak je funkčná závislosť rýchlosti nitridového leptania od týchto faktorov vyjadrená v kódovaných jednotkách, t.j. pracujeme s modelom, pomocou ktorého sme

testovali jednotlivé hypotézy, potom je potrebné pred dosadením do získaného predpisu transformovať úrovně faktorov pre ktoré chceme odozvu predikovať. Použijeme pritom rovnaké transformačné vzťahy, ktoré sme použili na začiatku experimentu. Predikovaná (modelovaná) hodnota musí vyjsť v oboch prípadoch rovná.

	A	B	C	D	E	F	G	H	I
1	Predikcia odozvy pre zvolene hodnoty vstupov:								
2	$y_L(t_1, t_3) =$	776,0625	+	-50,8125	*t ₁ +	153,0625	*t ₃ +	-76,8125	*t ₁ *t ₃
3									
4	$x_1=0,9 \rightarrow t_1 =$	$=(0,9-1)/0,2$		776,0625	1				
5	$x_2=150 \rightarrow t_2 =$	$=(150-162,5)/37,5$		-50,8125	=B4				
6	$x_3=280 \rightarrow t_3 =$	$=(280-300)/25$		153,0625	=B6				
7	$y_L(-0,5; -0,8) =$	$=D4+D5*B4+D6*B6+D7*B4*B6$		-76,8125	=B4*B6				
8					=SUMPRODUCT(D4:D7;				
9	$y_L(x_1, x_3) =$	-5415,375	+	4354,6875	*x ₁ +	21,485	*x ₃ +	-15,3625	*x ₁ *x ₃
10	$y_L(0,9; 280) =$	$=B9+D9*0,9+F9*280+H9*0,9*280$							
11									
12		=B9	1						
13		=D9	0,9						
14		=F9	280						
15		=H9	=C13*C14						
16			=SUMPRODUCT(B12:B15;C12:C15)						

1	Predikcia odozvy pre zvolene hodnoty vstupov:								
2	$y_L(t_1, t_3) =$	776,0625	+	-50,8125	*t ₁ +	153,063	*t ₃ +	-76,8125	*t ₁ *t ₃
3									
4	$x_1=0,9 \rightarrow t_1 =$	-0,5		776,0625	1				
5	$x_2=150 \rightarrow t_2 =$	-0,3333		-50,8125	-0,5000				
6	$x_3=280 \rightarrow t_3 =$	-0,8000		153,0625	-0,8000				
7	$y_L(-0,5; -0,8) =$	648,29375		-76,8125	0,4				
8					648,29375				
9	$y_L(x_1, x_3) =$	-5415,375	+	4354,688	*x ₁ +	21,485	*x ₃ +	-15,3625	*x ₁ *x ₃
10	$y_L(0,9; 280) =$	648,29375							
11									
12		-5415,375	1						
13		4354,6875	0,9						
14		21,485	280						
15		-15,3625	252						
16			648,29375						

Obr. 1.69 Predikcia odozvy pre zvolené hodnoty úrovní parametrov

Optimalizácia

Pri hľadaní optimálneho riešiteľa v Exceli používame jeho doplnok *Riešiteľ* (Solver), ktorý je potrebné pred prvým použitím spustiť rovnako ako *Analytické nástroje*. Následne tento doplnok otvárame postupnosťou krokov *Údaje* → *Riešiteľ*. Pred samotným aktivovaním dialógového okna *Riešiteľa* musia byť pripravené údaje, ktoré súvisia s výpočtom predikovanej hodnoty, ktorú chceme optimalizovať. Na výpočet predikovanej hodnoty je možné použiť

funkciu =SUMPRODUCT, ktorá počíta skalárny súčin vektora koeficientov nami nájdeného regresného modelu a vektora úrovní vstupných faktorov, ktoré vystupujú v predpise regresného modelu. Pri optimalizácii sme použili transformovaný regresný model $(\hat{f}_L)_t(t_1, t_3) = 776,0625 - 50,8125t_1 + 153,0625t_3 - 76,8125t_1t_3$, preto úrovně faktorov t_1 a t_3 musia byť vyjadrené v kódovaných hodnotách, viď obr. 1.70.

	A	B	C	D	E	F	G	H	I
17	koeficienty regresného modelu								
18	Optimalizácia v Exceli - hľadáme požadovanú hodnotu odozvy, napr. $800 \cdot 10^{(-10)}$ m/min - riešenie cez Solver								
19	Intercept	776,0625	1						
20	t1	-50,8125							
21	t3	153,0625							
22	t1*t3	-76,8125	=C20*C21						
23	y modelovane=		776,0625						

Obr. 1.70 Výpočet predikovanej hodnoty odozvy pre zvolené úrovně faktorov pred spustením dialógového okna Riešiteľa

V dialógovom okne *Riešiteľa* (obr.1.71) určíme cieľovú bunku, ktorý chceme optimalizovať, identifikujeme bunky s premennými, ktoré sa budú meniť, aby sa dosiahol cieľ, môžeme určiť obmedzenia a metódu riešenia.

Ako cieľová bunka je vybraná predikovaná hodnota

Ako cieľovú hodnotu si možno zvolit' maximum, minimum alebo presnú hodnotu

Obmedzenia určujúce nájdenie požadovanej hodnoty vstupov v rámci skúmaného priestoru

The screenshot shows the 'Parametre doplnku Riešiteľ' (Solver Parameters) dialog box. The 'Nastaviť cieľ:' (Set Objective) field is set to '\$C\$23'. The 'Do:' (To) section has three radio buttons: 'Maximum', 'Minimum', and 'Hodnota:' (Value Of), with 'Hodnota:' selected and the value '800' entered. The 'Zmenou premenných buniek:' (By Changing Variable Cells) field is set to '\$C\$20:\$C\$21'. The 'Podlieha obmedzeniam:' (Subject to the Constraints) section shows a list of constraints: '\$C\$20 <= 1', '\$C\$20 >= -1', '\$C\$21 <= 1', and '\$C\$21 >= -1'. The 'Vytvorte nezápornú hodnotu premenných, ktoré sú bez obmedzení' (Make Unconstrained Variables Non-Negative) checkbox is checked. The 'Vybrať metódu riešenia:' (Select a GRG Nonlinear engine) dropdown is set to 'Nelineárny algoritmus GRG'. The 'Metóda riešenia' (Solver Method) section contains text about selecting the appropriate algorithm. At the bottom, there are buttons for 'Pomocník' (Help), 'Riešiť' (Solve), and 'Zavrieť' (Close).

Hodnota 800 je zvolená ako cieľová hodnota

Identifikácia buniek s premennými, ktoré sa budú meniť

Tlačidlo na spustenie hľadania optimálneho riešenia

Obr. 1.71 Dialógové okno Riešiteľa

Excel našiel optimálne riešenie s hodnotou odozvy 799,9999, na obr. 1.72.

	A	B	C	D
17				
18	Optimalizacia v Exceli - hľadame požadovanú hodnotu odozvy, napr. $800 \cdot 10^{(-10)}$ m/min - riešenie cez Solver			
19	Intercept	776,0625	1	
20	t1	-50,8125	-0,0458146612192204	
21	t3	153,0625	0,138007504835749	
22	t1*t3	-76,8125	=C20*C21	
23	y modelovane=		=SUMPRODUCT(B19:B22;C19:C22)	- je požadovaná hodnota odozvy

Optimalizacia v Exceli - hľadame požadovanú hodnotu odozvy, napr. $800 \cdot 10^{(-10)}$ m/min - riešenie cez Solver			
Intercept	776,0625		
t1	-50,8125	-0,04581	
t3	153,0625	0,138008	
t1*t3	-76,8125	-0,00632	
y modelovane=		799,9999	- je požadovaná hodnota odozvy

Obr. 1.72 Nájdené riešenie

Záver:

Hľadali sme úrovně faktorov, na ktoré máme A_{medzera} medzi elektródami a $C_{\text{výkon}}$ na katóde nastaviť, aby rýchlosť nitrídového leptania dosahovala požadovanú hodnotu. Zvolili sme rýchlosť nitrídového leptania $800 \cdot 10^{-10}$ m/min. Podľa nájdeného riešenia má byť A_{medzera} medzi elektródami nastavená na hodnotu -0,046 a $C_{\text{výkon}}$ na katóde na 0,138. Vzhľadom na to, že toto sú kódované úrovně faktorov, spätnou transformáciou vieme vyjadriť úrovně faktorov v pôvodných jednotkách, t.j. faktor A_{medzera} medzi elektródami je potrebné nastaviť na 0,991 cm a faktor $C_{\text{výkon}}$ na 103,45 W. Riešenie je obr. 1.73.

	A	B	C	D	E	F	G	H
25	úroveň faktora v kodovaných jednotkách				úroveň faktora v pôvodných jednotkách			
26	t1		-0,045814849		x1	=0,2*C26+1		
27	t3		0,138008077		x2	=25*C27+100		

	A	B	C	D	E	F	G	H
25	úroveň faktora v kodovaných jednotkách				úroveň faktora v pôvodných jednotkách			
26	t1		-0,045814849		x1	0,99084		
27	t3		0,138008077		x2	103,45		

Obr. 1.73 Spätná transformácia umožňuje výpočet úrovni faktorov v pôvodných jednotkách