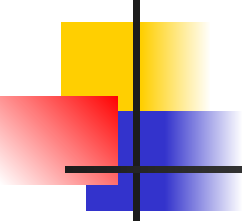


ОБРАБОТКА И АНАЛИЗ ЧИСЛОВОЙ ИНФОРМАЦИИ

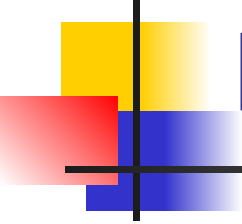
КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

©Румянцев Михаил Игоревич,
профессор, канд. техн. наук

Магнитогорск, 2007-2011



Корреляционный анализ – это метод математической статистики, который позволяет определить степень взаимосвязи между различными параметрами



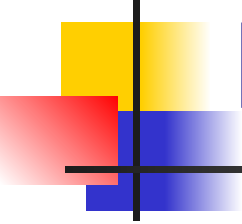
РАЗНОВИДНОСТИ КОРРЕЛЯЦИОННОГО АНАЛИЗА

ПАРНЫЙ

Оценивается степень взаимосвязи отклика Y и одного фактора X

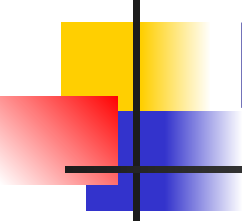
МНОЖЕСТВЕННЫЙ

Оценивается степень взаимосвязи отклика Y и нескольких факторов $X_1, \dots, X_j, \dots, X_m$



ХАРАКТЕРИСТИКА СТЕПЕНИ ВЗАИМОСВЯЗИ ПАРАМЕТРОВ

Характеристикой степени взаимосвязи параметров является статистическая величина, называемая коэффициентом корреляции

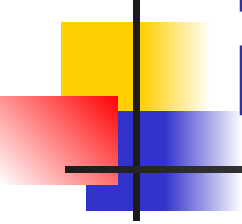


КОЭФФИЦИЕНТ ПАРНОЙ КОРРЕЛЯЦИИ

$$\rho = \frac{M \{ [x - M(X)] [y - M(Y)] \}}{\sqrt{M [x - M(X)]^2} \sqrt{M [y - M(Y)]^2}} = \frac{K_{XY}}{\sqrt{D(X)D(Y)}}$$

- K_{XY} - корреляционный момент. Он представляет собой математическое ожидание произведения отклонений значений x и y случайных величин X и Y от их математических ожиданий $M(X)$ и $M(Y)$;
- $D(X)$ - дисперсия случайной величины X ;
- $D(Y)$ - дисперсия случайной величины Y .

ВЫБОРОЧНАЯ ОЦЕНКА КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ


$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1) s_X s_Y}$$

$\bar{x}$ и $\bar{y}$ - средние выборочные значения фактора и отклика;

s_X и s_Y - выборочные стандартные отклонения фактора и отклика;

n - число наблюдений.

МАТРИЦА КОРРЕЛЯЦИИ

Таблица коэффициентов
парной корреляции,
которые отображают
взаимодействия отклика с
каждым из факторов
а также факторов между
собой

	Y	X_1	...	X_j	...	X_m
Y	1	r_{Y,X_1}	...	r_{Y,X_j}	...	r_{Y,X_m}
X_1	r_{Y,X_1}	1	...	r_{X_1,X_j}	...	r_{X_1,X_m}
...	...	...	1	...	...	...
X_j	r_{Y,X_j}	r_{X_1,X_j}	...	1	...	r_{X_j,X_m}
...	...	...	...	...	1	...
X_m	r_{Y,X_m}	r_{X_1,X_m}	...	r_{X_j,X_m}	...	1

СТРУКТУРА МАТРИЦЫ КОРРЕЛЯЦИИ

Коэффициенты парной корреляции отклика

Коэффициенты
корреляции
факторов

Главная
диагональ

	Y	X_1	...	X_j	...	X_m
Y	1	r_{Y,X_1}	...	r_{Y,X_j}	...	r_{Y,X_m}
X_1	r_{Y,X_1}	1	...	r_{X_1,X_j}	...	r_{X_1,X_m}
...	...	...	1	...	...	...
X_j	r_{Y,X_j}	r_{X_1,X_j}	...	1	...	r_{X_j,X_m}
...	...	...	...	...	1	...
X_m	r_{Y,X_m}	r_{X_1,X_m}	...	r_{X_j,X_m}	...	1

СИММЕТРИЧНОСТЬ МАТРИЦЫ КОРРЕЛЯЦИИ

Матрица
корреляции
симметрична
относительно
главной
диагонали

	Y	X_1	...	X_j	...	X_m
Y	1	r_{Y,X_1}	...	r_{Y,X_j}	...	r_{Y,X_m}
X_1	r_{Y,X_1}	1	...	r_{X_1,X_j}	...	r_{X_1,X_m}
...	...	...	1	...	...	...
X_j	r_{Y,X_j}	r_{X_1,X_j}	...	1	...	r_{X_j,X_m}
...	...	...	...	...	1	...
X_m	r_{Y,X_m}	r_{X_1,X_m}	...	r_{X_j,X_m}	...	1

КОЭФФИЦИЕНТ МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ

$$R = \sqrt{1 - \frac{\Delta_{YX}}{\Delta_{XX}}}$$

ОПРЕДЕЛИТЕЛЬ ВСЕЙ МАТРИЦЫ КОРРЕЛЯЦИИ

ОПРЕДЕЛИТЕЛЬ
МАТРИЦЫ
КОЭФФИЦИЕНТОВ
КОРРЕЛЯЦИИ
МЕЖДУ ФАКТОРАМИ

	Y	X_1	...	X_j	...	X_m
Y	1	r_{Y,X_1}	...	r_{Y,X_j}	...	r_{Y,X_m}
X_1	r_{Y,X_1}	1	...	r_{X_1,X_j}	...	r_{X_1,X_m}
...	...	...	1	...	...	...
X_j	r_{Y,X_j}	r_{X_1,X_j}	...	1	...	r_{X_j,X_m}
...	...	...	...	...	1	...
X_m	r_{Y,X_m}	r_{X_1,X_m}	...	r_{X_j,X_m}	...	1

СВОЙСТВА

КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

- Коэффициент корреляции не имеет размерности и поэтому сопоставим для различных статистических рядов.
- Значение коэффициента корреляции лежит в пределах от -1 до $+1$.
- Если коэффициент корреляции равен 1 , между параметрами существует функциональная зависимость.
- Коэффициент корреляции должен быть проверен на значимость.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

К
о
э
ф
ф
и
ц
и
е
н
т

К
о



Коэффициент
корреляции
действительно
не равен нулю?

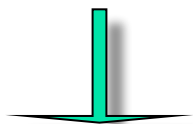
УСЛОВИЯ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ

Тест Стьюдента

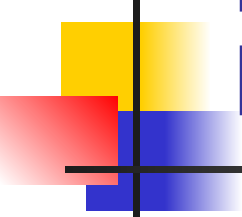
$$t = \frac{|r|}{\sqrt{1-r^2}} \sqrt{n-2} > t[\alpha; n-2]$$

t – рассчитанное число
Стьюдента

$t[\alpha; n-2]$ – табличное
число Стьюдента


$$r > r_{min} = \sqrt{\frac{1}{1 + \frac{n-2}{(t[\alpha; n-2])^2}}}$$

r_{min} – минимальное статистически значимое
значение коэффициента корреляции при
доверительной вероятности $p=1-\alpha$



УСЛОВИЕ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ

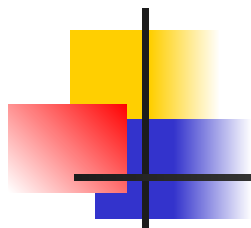
Тест Фишера

$$F_p = \frac{R^2}{(1 - R^2)} \frac{(n - m - 2)}{m} > F[\alpha; m; n - m - 2]$$

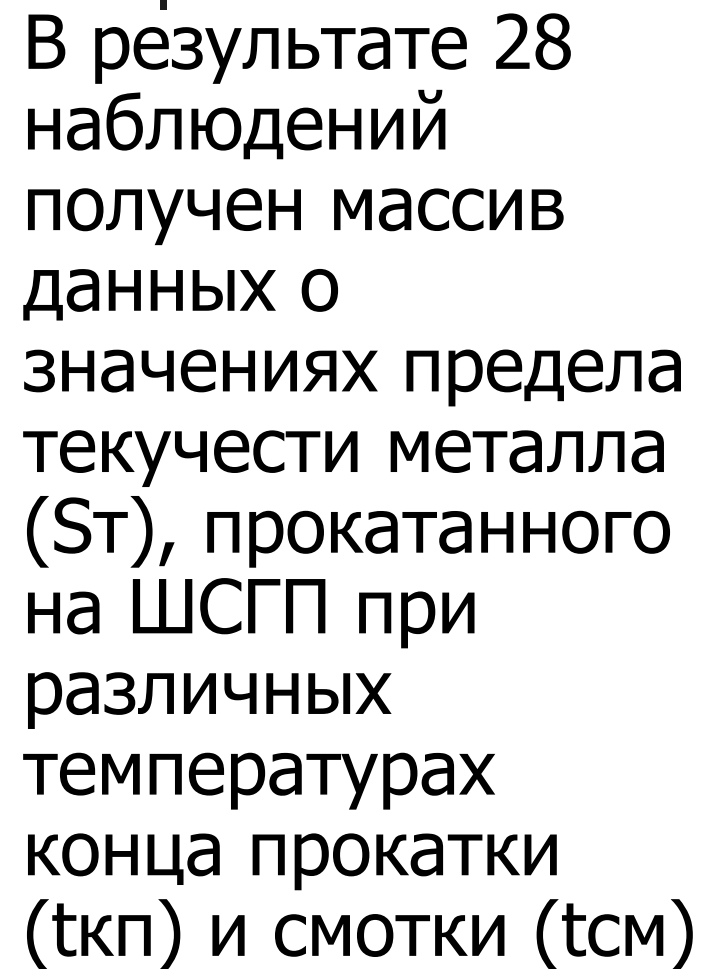
m – число факторов;

F_p – рассчитанное число Фишера;

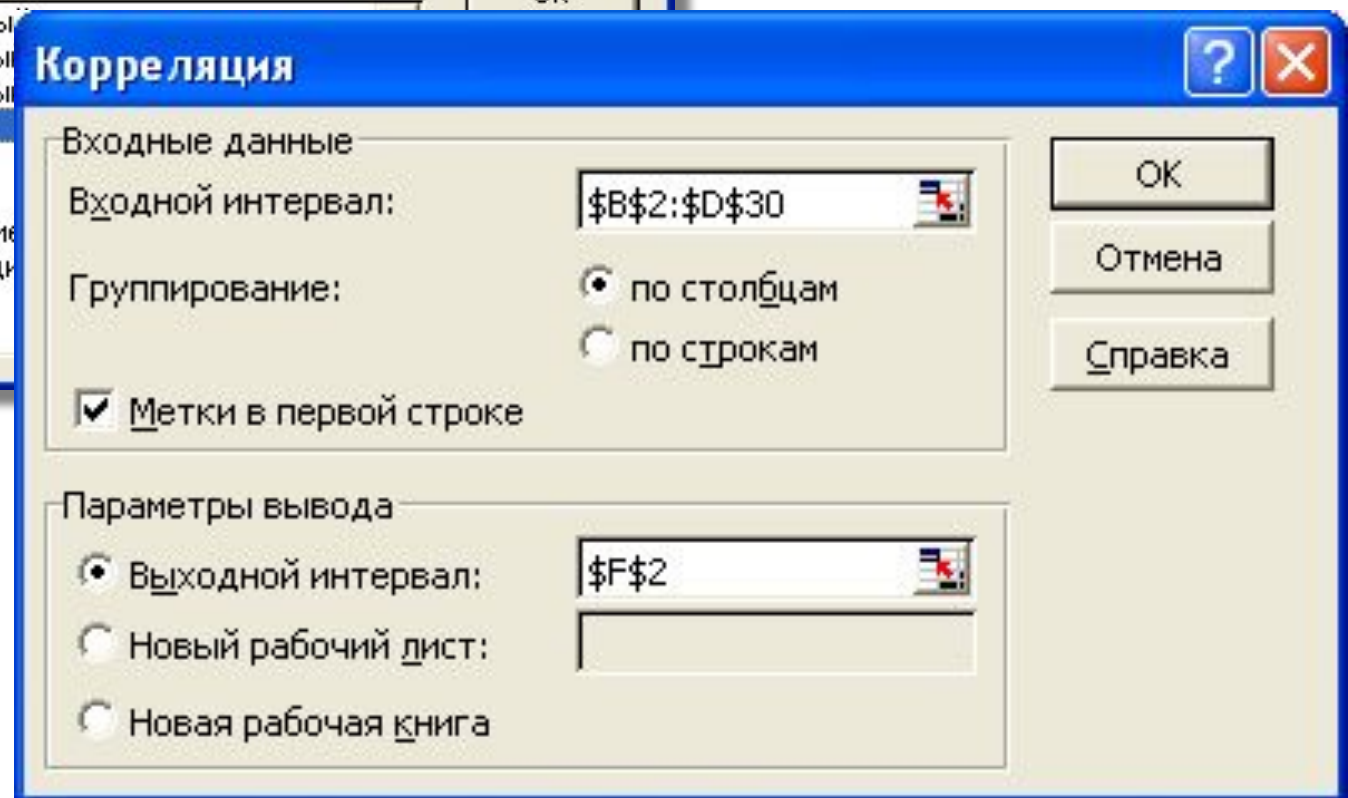
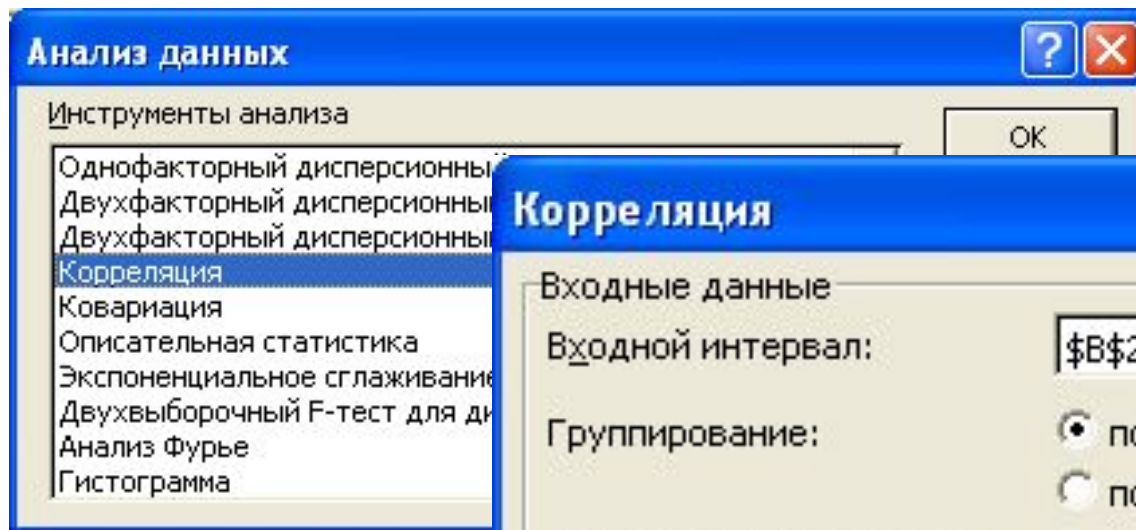
$F[\alpha; m; n - m - 2]$ – табличное число Фишера при доверительной вероятности $p = 1 - \alpha$.



ПРИМЕР КОРРЕЛЯЦИОННОГО АНАЛИЗА В MS EXCEL

31

ПРИМЕНЕНИЕ ИНСТРУМЕНТА «КОРРЕЛЯЦИЯ»



Коэффициент
парной корреляции
между $t_{кп}$ и $t_{см}$
 $r(t_{кп}; t_{см})$

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ ФУНКЦИЙ

Матрица корреляции $r(Y, X_i)$			
	Ст, МПа	$t_{кп}, C$	$t_{см}, C$
Ст, МПа	1	-0,474	-0,809
$t_{кп}, C$	-0,474	1	0,000
$t_{см}, C$	-0,809	0,000	1

$$r(t_{cm}; t_{kn}) = \text{КОРРЕЛ}(D3:D30; C3:C30)$$

Матрица корреляции действительно симметрична
относительно главной диагонали

ОЦЕНИВАНИЕ КОЭФФИЦИЕНТОВ ПАРНОЙ КОРРЕЛЯЦИИ

	A	B	C	D	E	F	G	H	I
1		ИСХОДНЫЕ ДАННЫЕ				Результат "КОРРЕЛЯЦИЯ"			
2		Ст, МПа	tkп, С	tcm, С			Sm, МПа	tkп, С	tcm, С
3		367,0	788	560		Ст, МПа	1		
4		241,2	788	584		tkп, С	-0,4736	1	
5		328,4	788	611		tcm, С	-0,8089	0	1
6		284,1	788	639					
7		330,9	788	657		Матрица корреляции r(Y,Xj)			
8		281,9	788	715			Ст, МПа	tkп, С	tcm, С
9		265,9	788	728		Ст, МПа	1	-0,474	-0,809
10		336,4	834	560		tkп, С	-0,474	1	0,000
11		320,5	834	584		tcm, С	-0,809	0,000	1
12		336,8	834	611					
13		299,4	834	639		Оценивание значимости коэффициентов корреляции			
14		269,3	834	657		n		28	
15		240,3	834	715		m		2	
16		206,3	834	728		p		0,95	
17		375	560			t[α ; n-2]		2,056	
18		375	584			t(St, tkп)		2,742	ДА
19		318,5	875	611				7,015	ДА
20								0,000	НЕТ
21									
22		288,1	875	715		АЯ КОРРЕЛЯЦИЯ			
23								0,937	ДА
24								86,80195	
25								0,95	
26								3,4028	
27								87,9	%
28		270,3	917	657					
29			917	715					
30			917	728					
31									

=СЧЁТ(B3:B30)

=СЧЁТЗ(C2:D2)

С клавиатуры

=СТЮДРАСПОБР(1-H17;H15-2)

=ABS(G10)*КОРЕНЬ(\$H\$15-2)/КОРЕНЬ(1-G10^2)

=ABS(G11)*КОРЕНЬ(\$H\$15-2)/КОРЕНЬ(1-G11^2)

=ABS(H11)*КОРЕНЬ(\$H\$15-2)/КОРЕНЬ(1-H11^2)

=ЕСЛИ(H19>\$H\$18;"ДА";"НЕТ")

РАСЧЕТ И ОЦЕНИВАНИЕ

	A	B	C	D	E	F	G	H	I
	Исходные данные				Результат "КОРРЕЛЯЦИЯ"				
2	Ст, МПа	ткп,С	тсм,С		Ст, МПа	ткп,С	тсм,С		
3	364,0	788	560		Ст, МПа	1			
4	341,2	788	584		ткп,С	-0,4736	1		
5	336,4	788	611		тсм,С	-0,8089	0,000	1	
6	284,1	788	639						
7	330,9	788	657		Матрица корреляции r(Y,Xj)				
8	281,9	788	715		Ст, МПа	ткп,С	тсм,С		
9	265,9	788	728		Ст, МПа	1	-0,474	-0,809	
10	336,4	834	560		ткп,С	-0,474	1	0,000	
11	320,5	834	584		тсм,С	-0,809	0,000	1	
12	336,8	834	611						

=КОРЕНЬ(1-МОПРЕД(G9:I11)/МОПРЕД(H10:I11))

значимости
рв корреляции

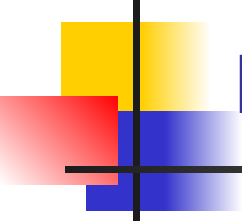
$$=H24*H24*(H15-H16-2)/((1-H24*H24)*H16)$$

С клавиатуры

$$= \text{FRASПОВР}(1-H26; H16; H15-H16-2)$$
~~$$= 100 * H24 * H24$$~~

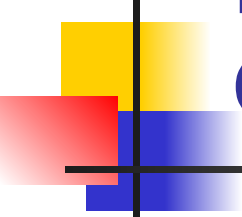
=ЕСЛИ(Н25>Н27;"ДА";"НЕТ")

9:I11)/МОПРЕД(Н10:I11))					значимости ов корреляции	
15	240,3	834	715	n	28	
/((1-Н24*Н24)*Н16)					2	
					0,95	
18	290,3	875	584	t[α ;n-2]	2,056	
19	318,5	875	611	t(St,tkn)	2,742	ДА
20	283,2	875	639	t(St,tcm)	7,015	ДА
21	281,0	875	657	t(tcm,tkn)	0,000	НЕТ
,Н15-Н16-2)				715		
				728		
МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ						
24	304,4	917	560	R	0,937	ДА
25	307,5	917	584	Fp	86,80195	
26	268,8	917	611	p	0,95	
27	225,4	917	639	F[α ;m;n-m-2]	3,4028	
28	270,3	917	657	D	87,9	%
29	181,3	917	715			
;"НЕТ")				917	728	



Связь между какими величинами анализировалась?

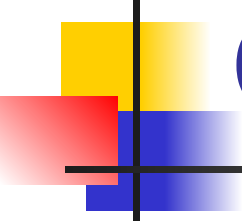
Анализировалась связь между пределом текучести металла σ_T , температурой конца прокатки $t_{кп}$ и смотки $t_{см}$ при прокатке на ШСГП.



Какие коэффициенты парной корреляции являются статистически значимыми?

С доверительной вероятностью 95% статистически значимыми являются коэффициенты корреляции между пределом текучести и температурой конца прокатки $r(\sigma_T ; t_{кп}) = -0,474$ а также между пределом текучести и температурой смотки $r(\sigma_T ; t_{см}) = -0,809$.

Значимость коэффициентов подтверждается тем, что соответствующие расчетные числа Стьюдента $t(\sigma_T ; t_{кп}) = 2,742$ и $t(\sigma_T ; t_{см}) = 7,015$ больше табличного $t[0,05; 26] = 2,056$.

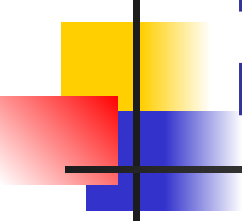


О чем это свидетельствует?

Следовательно, предел текучести металла, прокатанного на ШСГП, связан с температурой конца прокатки и смотки.

Так как коэффициенты корреляции отрицательные, увеличение как температуры прокатки, так и температуры смотки уменьшает предел текучести прокатанного металла.

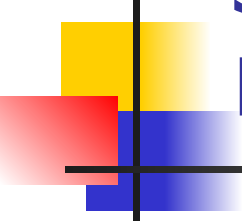
Так как $|r(\sigma_T ; t_{см})| > |r(\sigma_T ; t_{кп})|$, степень влияния температуры смотки больше чем температуры конца прокатки.



Является ли значимым коэффициент множественной корреляции? Что это означает?

С доверительной вероятностью 95% коэффициент множественной корреляции $R(\sigma_T; t_{kp}; t_{cm}) = 0,937$ является статистически значимым, т. к. расчетное число Фишера $F_p = 86,802$ больше табличного $F[0,05; 2; 24] = 3,4028$.

Это означает, что предел текучести металла, прокатанного на ШСГП, обусловлен совместным действием температуры конца прокатки и смотки.



О чем свидетельствует значение коэффициента множественной детерминации?

Коэффициент множественной детерминации $D=0,879$ свидетельствует, что при прокатке на ШСГП предел текучести металла на 87,9% обусловлен сочетанием температуры конца прокатки и смотки.