

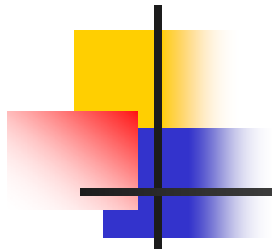


ОБРАБОТКА И АНАЛИЗ ЧИСЛОВОЙ ИНФОРМАЦИИ

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

© Румянцев Михаил Игоревич,
профессор, канд. техн. Наук
с Адаптивными правками С.А. Левандовского

© Каф. ТОМ, МГТУ им. Г.И. Носова



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

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

- 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**.
- **Положительное** значение коэффициента корреляции указывает на возрастание отклика с увеличением фактора. **Отрицательное** значение r свидетельствует об убывании Y при возрастании X .


СВОЙСТВА

КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

- Если коэффициент корреляции **равен 1**, между параметрами существует функциональная зависимость.
- Коэффициент корреляции должен быть проверен на значимость.
- Равенство коэффициента парной корреляции нулю не означает отсутствия связи между откликом и фактором. В литературе отмечается, что значение $r = 0$ указывает лишь что эта взаимосвязь не является линейной, но не опровергает возможность существования между ними иной, например экспоненциальной, зависимости.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Коэффициент
корреляции
вычисляется
на основании
выборочных
данных
и поэтому
является
случайной
величиной



**Коэффициент
корреляции
действительно
не равен нулю?**

УСЛОВИЯ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА ПАРНОЙ КОРРЕЛЯЦИИ

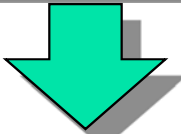
Тест
Стьюдента

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

t – рассчитанное число
Стьюдента

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

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


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

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

КОЭФФИЦИЕНТ МНОЖЕСТВЕННОЙ ДЕТЕРМИНАЦИИ

$$D = R^2$$

При анализе степени совместного влияния комплекса факторов на отклик часто используют коэффициент множественной детерминации **D**.

Его значение показывает, на сколько процентов изменчивость отклика обусловлена совместным действием рассматриваемых факторов.

УСЛОВИЕ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ

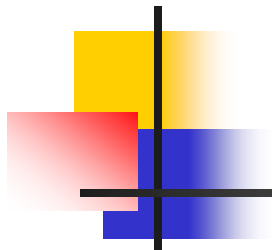
Тест Фишера

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

m – число факторов;

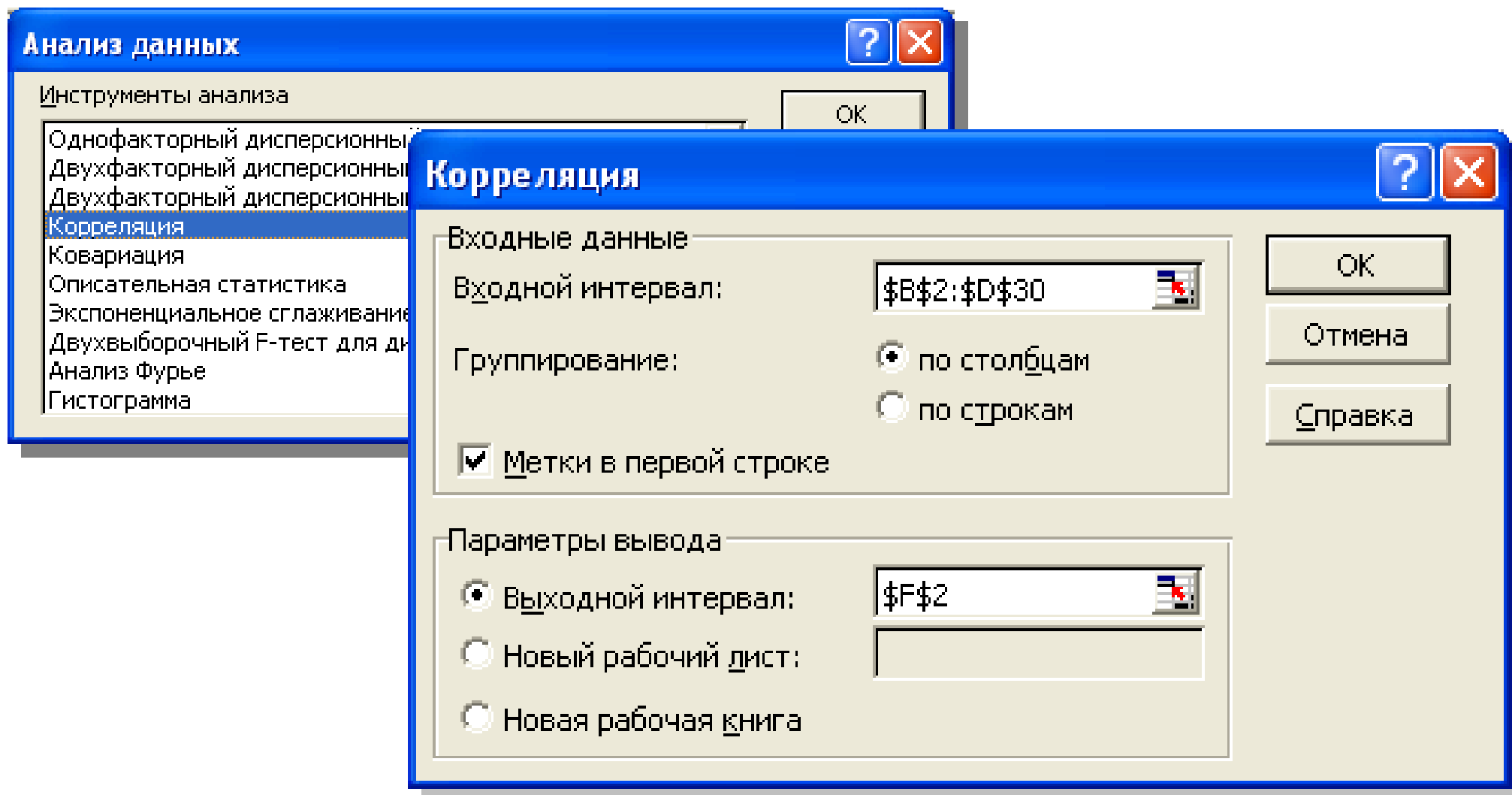
F_p – рассчитанное число Фишера;

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



ПРИМЕР КОРРЕЛЯЦИОННОГО АНАЛИЗА В MS EXCEL

ПРИМЕНЕНИЕ ИНСТРУМЕНТА «КОРРЕЛЯЦИЯ»



ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ ФУНКЦИЙ

	В	С	D	E	F	G	H	I
1	ИСХОДНЫЕ ДАННЫЕ				Результат "КОРРЕЛЯЦИЯ"			
2	σ_T , МПа	$t_{кп}$, С	$t_{сМ}$, С			S_m , МПа	$t_{кп}$, С	$t_{сМ}$, С
3	361,0	875	560		S_T , МПа	1		
4	341,2	788	584		$t_{кп}$, С	-0,4736	1	
5	338,4	788	611		$t_{сМ}$, С	-0,8089	0	1
6	284,1	788	639					
7	330,9	788	657		Матрица корреляции $r(Y, X_j)$			
8	281,9	788	715		S_T , МПа	$t_{кп}$, С	$t_{сМ}$, С	
9					S_T , МПа	1	-0,474	-0,809
10					$t_{кп}$, С	-0,474	1	0,000
11	320,5	834	584		$t_{сМ}$, С	-0,809	0,000	1
12								
13					Оценивание значимости коэффициентов корреляции			
14					n		23	
15					m		2	
16					p		0,95	
17	354,2	875	560		$r_{(T, t_{кп})}$		2,056	
18					$r_{(T, t_{сМ})}$		2,742	ДА
19	331,0	875	584		$r_{(T, t_{кп})}$		7,015	ДА
20	331,0	875	657		$r_{(T, t_{сМ})}$		0,000	НЕТ
21					КОЭФФИЦИЕНТ ПАРЦИАЛЬНОЙ КОРРЕЛЯЦИИ			
22					R		0,937	ДА
23					p		86,80195	
24	304,4	917	560				0,95	
25					F[α ; m; n-m-2]		3,4028	
26								
27	225,4	917	639					

$$r(\sigma_T ; t_{кп}) = \text{КОРРЕЛ}(B3:B30; C3:C30)$$

$$r(\sigma_T ; t_{сМ}) = \text{КОРРЕЛ}(B3:B30; D3:D30)$$

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

$$r(t_{кп}; \sigma_T) = \text{КОРРЕЛ}(C3:C30; B3:B30)$$

$$r(t_{сМ}; \sigma_T) = \text{КОРРЕЛ}(D3:D30; B3:B30)$$

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

Матрица корреляции действительно симметрична относительно главной диагонали

РАСЧЕТ И ОЦЕНИВАНИЕ КОЭФФИЦИЕНТА МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ

Исходные данные				Результат "КОРРЕЛЯЦИЯ"		
	Ст, МПа	tkп, С	tсм, С	Ст, МПа	tkп, С	tсм, С
3	364,0	788	560	Ст, МПа	1	
4	341,2	788	584	tkп, С	-0,4736	1
5	355,2	788	611	tсм, С	-0,809	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	tсм, С
9	265,9	788	728	Ст, МПа	1	-0,474
10	336,4	834	560	tkп, С	-0,474	1
11	320,5	834	584	tсм, С	-0,809	0,000
12	336,8	834	611			
15	240,3	834	715	n	28	
18	290,3	875	584	t[α ; n-2]	2,056	
19	318,5	875	611	t(St, tkп)	2,742	ДА
20	283,2	875	639	t(St, tсм)	7,015	ДА
21	281,0	875	657	t(tсм, tkп)	0,000	НЕТ
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			

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

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

С клавиатуры

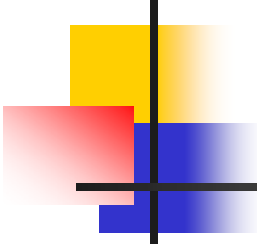
=ФРАСПОБР(1-H26;H16;H15-H16-2)

=100*H24*H24

=ЕСЛИ(H25>H27;"ДА";"НЕТ")

значимости
в корреляции

МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ



Связь между какими величинами анализировалась?

Анализировалась связь между пределом текучести металла σ_T , температурой конца прокатки **ткп** 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 $|r(\sigma_T ; t_{CM})| > |r(\sigma_T ; t_{KN})|$, 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 **$R(\sigma_T; t_{кп}; t_{см}) = 0,937$** 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.