

Лекция №3

**Корреляционный и статистический анализ
экспериментальных данных в программе
Excel**

§1 Корреляционная матрица

Корреляционным анализом называется совокупность методов обнаружения корреляционной зависимости между случайными величинами или признаками.

При большом числе наблюдений, когда коэффициенты корреляции необходимо последовательно вычислять для нескольких выборок, для удобства получаемые коэффициенты сводят в таблицы, называемые корреляционными

На основании значений матрицы и парных коэффициентов корреляции $r(x_j, x_k)$ могут делаться разнообразные выводы о природе связей между явлениями, зависящие от направления исследования и изучаемой предметной области.

Корреляционная матрица — это квадратная таблица, в которой на пересечении соответствующих строки и столбца находится коэффициент корреляции между соответствующими параметрами.

	1	2	3	4	5	6	7
1	$r=1$						
2	$r(1;2)$	$r=1$					
3	$r(1;3)$	$r(2;3)$	$r=1$				
4	$r(1;4)$	$r(2;4)$	$r(3;4)$	$r=1$			
5	$r(1;5)$	$r(2;5)$	$r(3;5)$	$r(4;5)$	$r=1$		
6	$r(1;6)$	$r(2;6)$	$r(3;6)$	$r(4;6)$	$r(5;6)$	$r=1$	
7	$r(1;7)$	$r(2;7)$	$r(3;7)$	$r(4;7)$	$r(5;7)$	$r(6;7)$	$r=1$

1, 2, ..., 7 – отдельные показатели;

r – коэффициент корреляции между показателями.

Коэффициентом корреляции называется величина, абсолютное значение которой оценивает тесноту взаимосвязи между двумя показателями.

Значение коэффициента корреляции будем интерпретировать следующим образом:

$0 < |r| \leq 0,2$ – связи нет;

$0,2 < |r| \leq 0,4$ – связь слабая;

$0,4 < |r| \leq 0,7$ – связь средняя;

$0,7 < |r| \leq 1$ – связь сильная;

$r < 0$ – связь обратная;

$r > 0$ – связь прямая.

Коэффициент корреляции находящийся на пересечении одного и того же показателя равен 1, так как каждый столбец во входном

Хотя в результате будет получена треугольная матрица, корреляционная матрица симметрична. Подразумевается, что в пустых клетках в правой верхней половине таблицы находятся те же коэффициенты корреляции, что и в нижней левой (симметрично расположенные относительно диагонали). То есть $r(x_j, x_k) = r(x_k, x_j)$.

Например: $r(2;5) = r(5;2)$

Этапы корреляционного анализа

1. Оформить в программе Excel экспериментальные данные в виде таблицы;
2. Построить корреляционную матрицу с помощью пакета **“Анализ данных...”**;
3. Оценить каждый коэффициент корреляции;
4. Нарисовать граф (графическую структуру) корреляционной матрицы средствами рисования программы Excel или Word;

§2 Построение корреляционной матрицы в Excel

В электронной таблице Excel для вычисления корреляционных матриц используется процедура **Корреляция** из пакета или надстройки **Анализ данных**.

Процедура позволяет получить корреляционную матрицу, содержащую коэффициенты корреляции между различными параметрами.

Для реализации процедуры необходимо установить надстройку «**Пакет анализа**», которая выведет команду **Анализ данных** в пункт **Данные** (см лекцию2).

Пусть спортивные результаты группы спортсменов набраны в программе Excel.

Тогда для построения корреляционной матрицы по показателям надо:

- 1) Выбрать пункт «Данные» → команду «Анализ данных» → инструмент «Корреляция» → ОК

Откроется диалоговое окно!

- 2) Для входного интервала на рабочем

	A	B	C	D	E	F	G
1							
2			Бег	Бег	Бег	Бег	
3			на 30 м	на 60 м	на 100м	на 200м	
4			(с)	(с)	(с)	(с)	
5			1	2	3	4	
6		Иванов	4	8,2	14,6	27,3	
7		Петров	3,8	8,3	14,7	27,5	
8		Николаев	3,7	8,5	14,8	27,6	
9		Новиков	3,9	8,8	14,5	27,9	
10		Попов	4,2	8,7	14,3	28	
11		Сидоров	4,1	8,9	14,2	28,2	
12		Мамин	4,5	8,3	14,1	28,1	
13		Шишкин	4,7	8	13,8	27,5	
14							

Корреляция

Входные данные

Входной интервал:

Группирование:

по столбцам

по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

ОК

Отмена

Справка

3) Установить галочку в окошке «Метки в первой строке» → Группирование по столбцам → Выходной интервал в пустой ячейке → ОК
Построится корреляционная матрица

Все получившиеся коэффициенты корреляции округлить до сотых, используя инструмент «Уменьшение разрядности»

	A	B	C	D	E	F	G
1							
2			Бег	Бег	Бег	Бег	
3			на 30 м	на 60 м	на 100м	на 200м	
4			(с)	(с)	(с)	(с)	
5			1	2	3	4	
6		Иванов	4	8,2	14,6	27,3	
7		Петров	3,8	8,3	14,7	27,5	
8		Николаев	3,7	8,5	14,8	27,6	
9		Новиков	3,9	8,8	14,5	27,9	
10		Попов	4,2	8,7	14,3	28	
11		Сидоров	4,1	8,9	14,2	28,2	
12		Мамин	4,5	8,3	14,1	28,1	
13		Шишкин	4,7	8	13,8	27,5	
14							

1: Корреляция

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

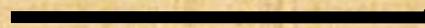
OK
Отмена
Справка

§3 Граф корреляционной матрицы

Граф построенный на основе корреляционной матрицы отображает наиболее значимые связи между спортивными результатами, а так же визуально выявляет не значимые показатели.

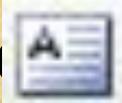
Построение графа корреляционной матрицы

1) Сначала вводятся обозначения:

 — сильные связи

 — связи средней силы

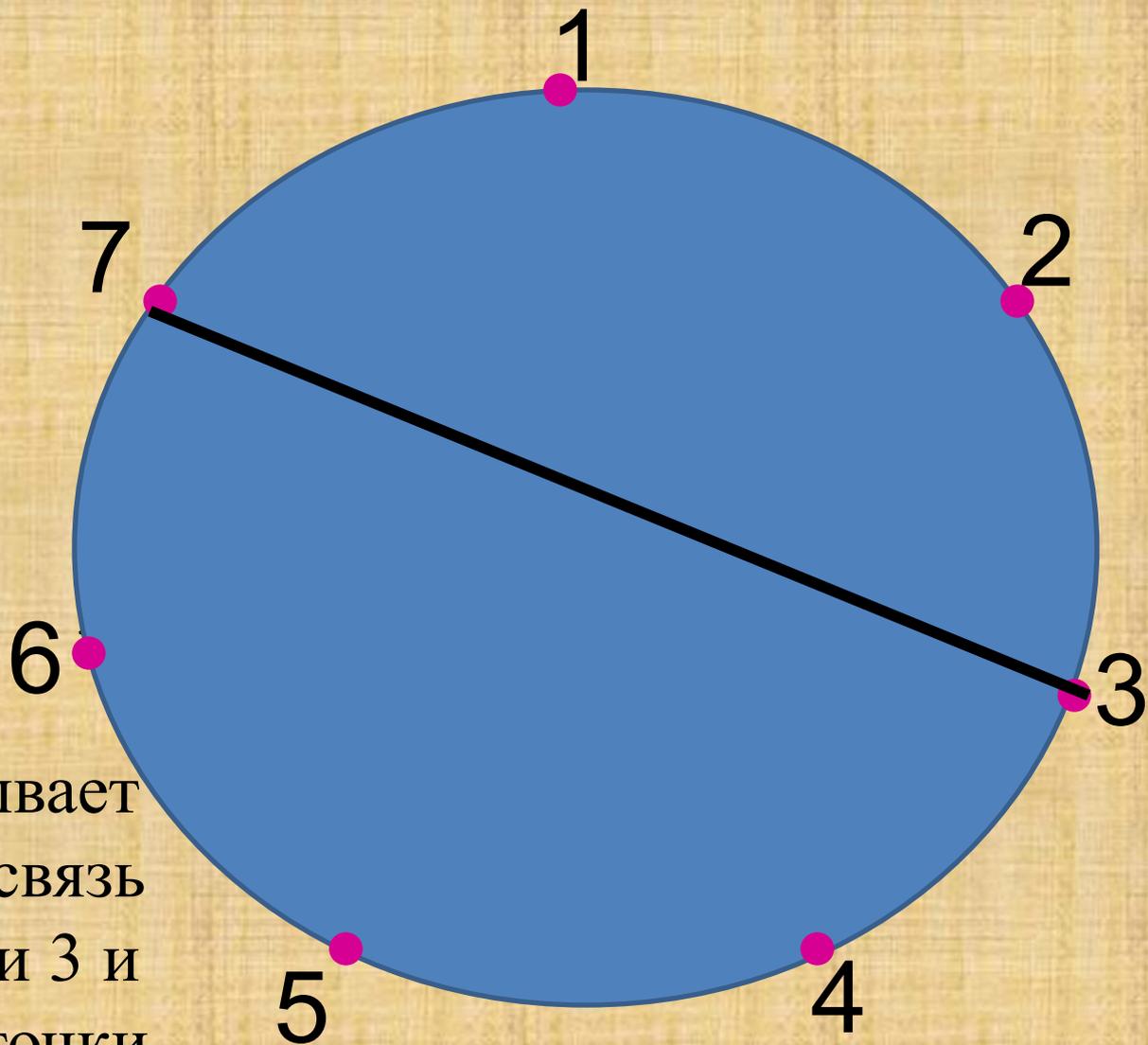
 — слабые связи

2) Далее с помощью инструмента  «надпись» обозначения показателей располагаются равномерно, по порядку, по кругу. Для этого удобно на окружности отметить нужное количество точек на одинаковом расстоянии. У каждой надписи **отменить** контур и заливку фигуры.

Например:

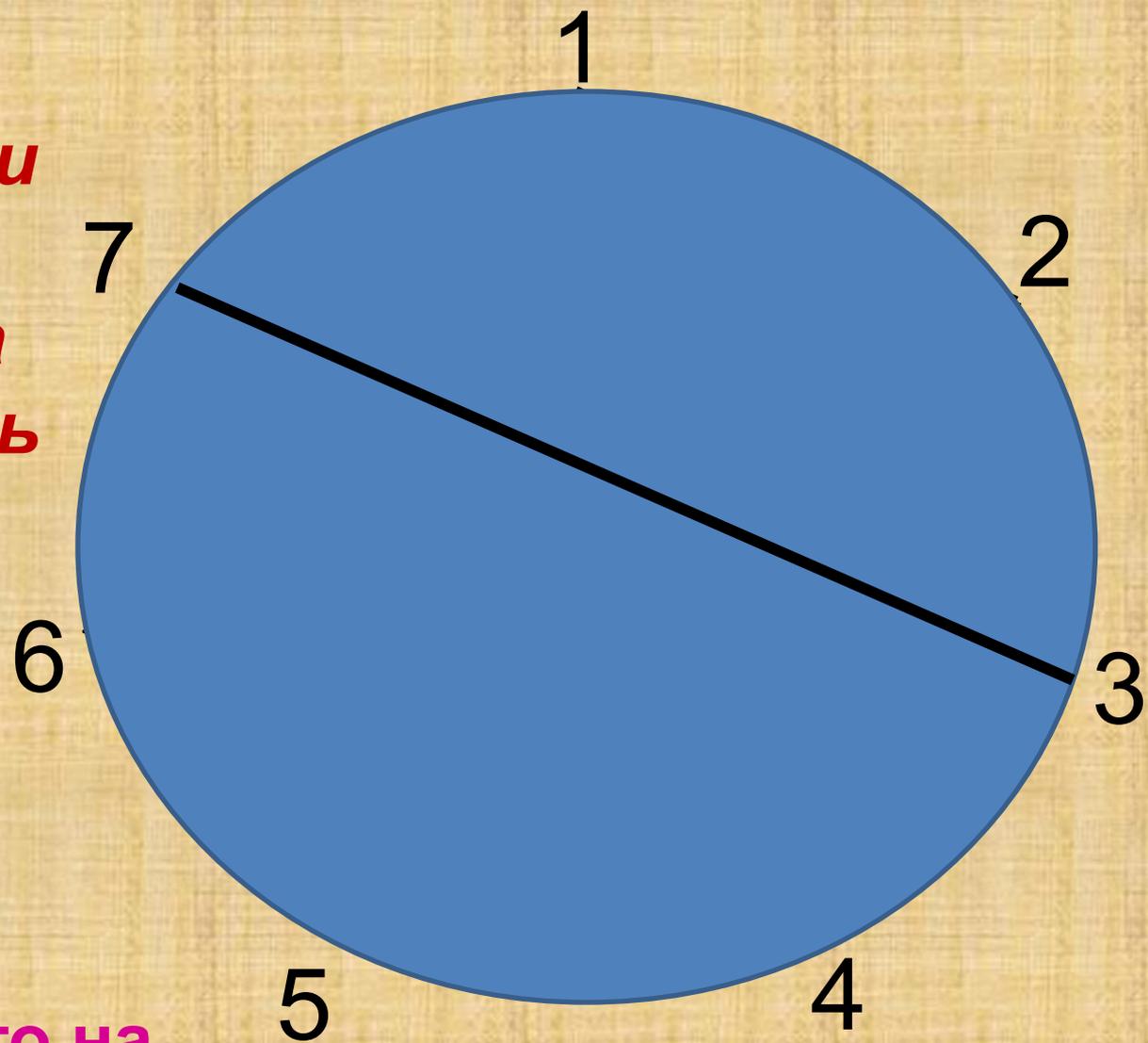
3) Соединить показатели линиями, применяя к ним нужный формат контура.

Если $r(3;7)$ указывает на сильную взаимосвязь между показателями 3 и 7, то от точки 3 до точки 7 рисуем линию и изменяем её толщину.



Причём не важно, какая это связь **прямая** или **обратная**.

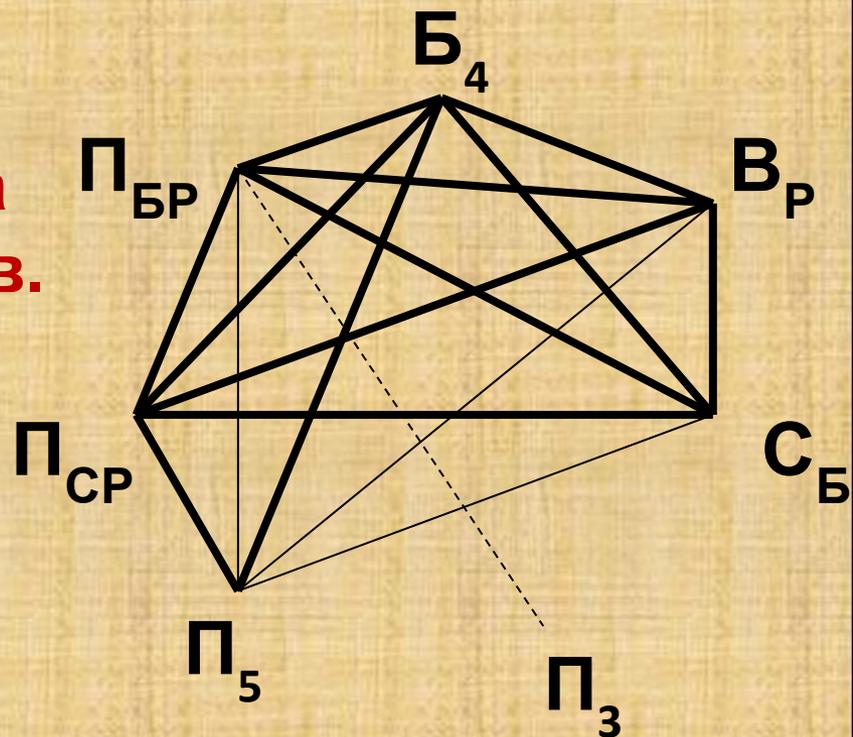
4) Когда все значимые связи будут изображены на графе, удалить круг (окружность)



Если $|r| \leq 0,2$, то на графике такую взаимосвязь отображать не нужно

Например:

Пусть показатель Π_3 - результаты тройного прыжка в данной группе спортсменов. Тогда из данного графа видно, что тройной прыжок выпадает из общей картины взаимосвязей.



Следовательно, в данной группе спортсменов тройной прыжок можно исключить из тренировочного процесса и заменить его другими упражнениями, которые повышают эффективность тренировок.

§4 Статистический анализ в Excel

Microsoft Excel содержит большое число встроенных *статистических функций*.

После установки надстройки «*Пакет анализа*» расширяются возможности обработки статистической информации, кроме того процесс статистического анализа значительно упрощается.

Средства, включенные в «*Пакет анализа*» позволяют выполнить проверку статистических гипотез о параметрах распределения случайных величин, построение линейных и нелинейных уравнений регрессии,

Одним из инструментов пакета «*Анализ данных...*» является инструмент «*Описательная статистика*».

Цель «Описательной статистики» - обобщить первичные результаты, полученные в результате наблюдений и экспериментов.

Это средство анализа служит для создания одномерного статистического отчёта содержащего информацию о центральной тенденции и изменчивости или вариации входных данных.

В состав описательной статистики входят следующие

ЦЕНТРАЛЬНЫЕ ТЕНДЕНЦИИ

*среднее;
стандартная
ошибка;
медиана;
мода*

ИЗМЕНЧИВОСТЬ ДАННЫХ

*стандартное
отклонение;
дисперсия выборки;
эксцесс;
асимметричность;
интервал;
минимум;
максимум;
сумма;
счет*

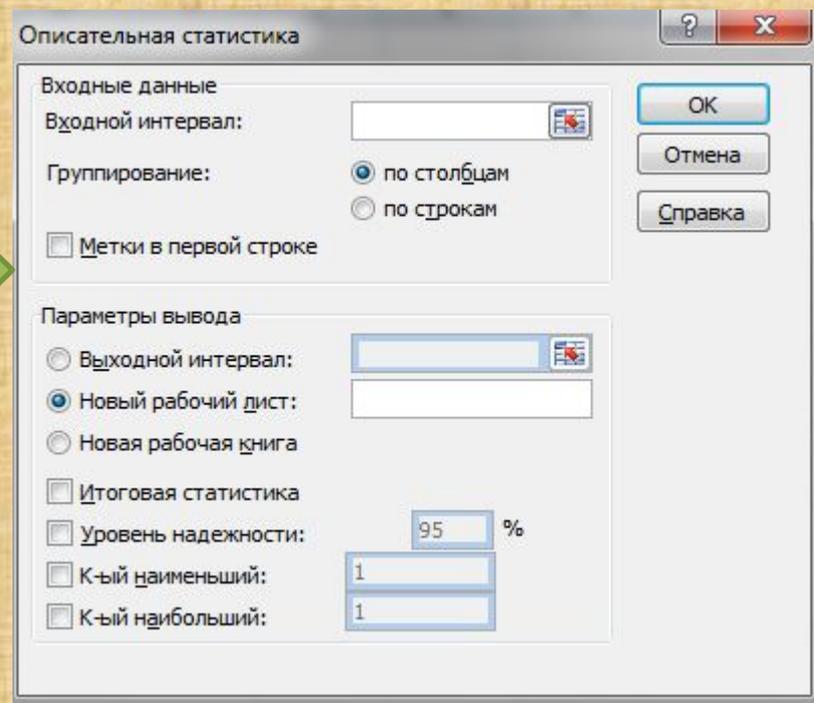
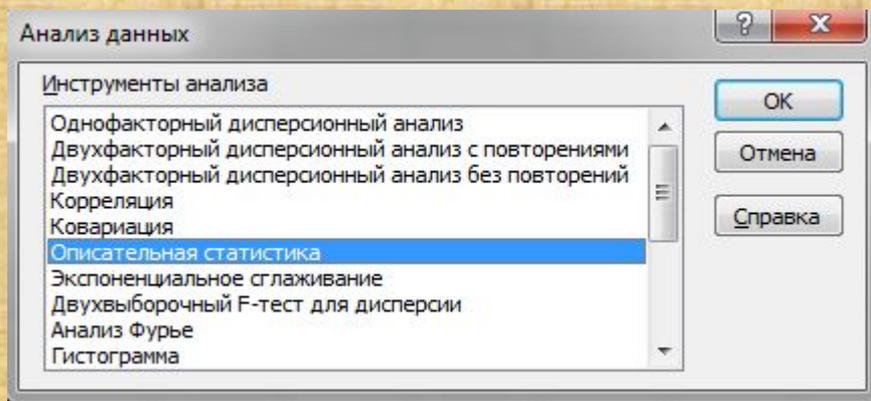
Порядок построения таблицы основных статистических характеристик

Пусть в программе Excel набраны
данные для построения таблицы с
итоговой описательной статистикой надо:

- 1) Выбрать пункт «**Данные**» → команду
«**Анализ данных**» → инструмент
«**Описательная статистика**» → ОК

Откроется диалоговое окно!

- 2) Для входного интервала на рабочем
листе выделить диапазон данных с
метками (названиями столбцов)



3) Установить галочку в окошке «Метки в первой строке» → Группирование по столбцам → Выходной интервал в пустой ячейке → Флажок «Итоговая статистика» → уровень надёжности 95% → ОК

	A	B	C	D	E	F	G
1							
2			Бег	Бег	Бег	Бег	
3			на 30 м	на 60 м	на 100м	на 200м	
4			(с)	(с)	(с)	(с)	
5			1	2	3	4	
6		Иванов	4	8,2	14,6	27,3	
7		Петров	3,8	8,3	14,7	27,5	
8		Николаев	3,7	8,5	14,8	27,6	
9		Новиков	3,9	8,8	14,5	27,9	
10		Попов	4,2	8,7	14,3	28	
11		Сидоров	4,1	8,9	14,2	28,2	
12		Мамин	4,5	8,3	14,1	28,1	
13		Шишкин	4,7	8	13,8	27,5	
14							



	H	I
	2	
	Среднее	8,4625
	Стандартная ошибка	0,111704
	Медиана	8,4
	Мода	8,3
	Стандартное отклонение	0,315945
	Дисперсия выборки	0,099821
	Эксцесс	-1,29767
	Асимметричность	0,060584
	Интервал	0,9
	Минимум	8
	Максимум	8,9
	Сумма	67,7
	Счет	8
	Уровень надежности(95,0%)	0,264137

Округлить характеристики, которые необходимо, до сотых с помощью значка на панели инструментов «Уменьшение разрядности».

Сделать вывод

Подготовить презентацию к
защите РГР «Корреляционный
и статистический анализ
данных»

Требования к презентации:

1. Презентация должна состоять минимум из 7-ми слайдов (титульный лист, исходные данные, корреляционная матрица, граф корреляционной матрицы, итоговая статистика одного показателя гистограмма и вывод)
2. Эффект смены каждого слайда.
3. Эффекты анимации (вход каждому объекту на слайдах, выделение отдельных элементов.)
4. Полная разгруппировка графа и постепенное рисование отдельных элементов (см.след. слайд).
5. Все эффекты по щелчку.
6. Поддержание единого стиля.

Порядок создания слайда:

1. Создать картинку, т.е. расположить все нужные объекты на слайде в нужных местах. Все лишние объекты удалять.
2. Установить фон слайда или оформление слайда
3. Установить смену слайда
4. Установить эффект входа объекту «**заголовок**» (Выделить заголовок → Анимация → Настройка анимации → Добавить эффект → Вход → выбрать эффект)
5. Установить эффекты входа остальным объектам слайда
6. При необходимости установить эффект выделения отдельному элементу (Выделить элемент → Добавить эффект → Выделение → выбрать эффект)

Графическое изображение корреляционной матрицы.

