

Практическое занятие №1 «Корреляционный анализ»

по дисциплине «Многомерный
статистический анализ в
социологических
исследованиях»

План занятия

1. Корреляционный анализ.
2. Проверка гипотез о взаимосвязи переменных.

Примеры

1. **Менеджер** интересуется, зависит ли объем продаж в этом месяце от объема рекламы в этом же периоде?
2. **Преподаватель** хочет выяснить, есть ли зависимость между количеством часов, потраченных студентом на занятия, и результатами экзамена?
3. **Врач** исследует, влияет ли кофеин на сердечные болезни и существует ли связь между возрастом человека и его кровяным давлением?
4. **Социолог** исследует, какова связь между уровнем преступности и уровнем безработицы в регионе? Есть ли зависимость между расходами на жилье и совокупным доходом семьи? Связаны ли доход от профессиональной деятельности и продолжительность образования?

Постановка проблемы

- Наша цель – научиться отвечать на четыре вопроса:
- Вопрос 1. **Существует ли связь** между двумя или более переменными?
- Вопрос 2. Какой **тип** имеет эта связь?
- Вопрос 3. Насколько она **сильна**?
- Вопрос 4. Какой можно сделать **прогноз**, основываясь на этой связи?

Корреляционный анализ

- это совокупность основанных на математической теории корреляции методов обнаружения корреляционной зависимости между двумя случайными признаками или факторами
- это проверка гипотез о связях между переменными с использованием коэффициентов корреляции
- «Оба термина, — пишет Е.В. Сидоренко, — **корреляционная связь и корреляционная зависимость** — часто используются как синонимы. Зависимость подразумевает влияние, связь — любые согласованные изменения, которые могут объясняться сотнями

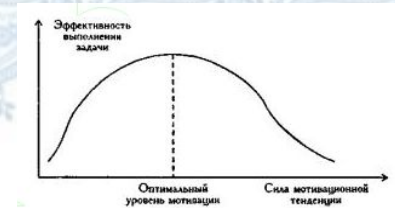
Корреляция и причинная связь

- Когда проверка гипотезы показывает, что существует значимая линейная связь между переменными, исследователь должен рассмотреть возможные виды связи между переменными и выбрать ту, диктуется логикой исследова



Корреляционные связи

- **По форме** корреляционная связь может быть прямолинейной или криволинейной
- **По направлению** корреляционная связь может быть положительной ("прямой") и отрицательной ("обратной")
- **Степень, сила или теснота** корреляционной связи определяется по величине коэффициента корреляции



Корреляционные связи

- *положительная корреляция* (большие значения одного набора связаны с большими значениями другого)
- *отрицательная корреляция* (малые значения одного набора связаны с большими значениями другого)
- *нулевая корреляция* (данные двух наборов никак не связаны)

- Величина (сила) связи и ее значимость (достоверность) представляют две различные характеристики связи. В общем случае, чем сильнее связь, тем более значимой она является
- Нулевая гипотеза утверждает, что для генеральной совокупности, из которой была извлечена выборка, связь между переменными полностью отсутствует, т.е. значение проверяемого показателя меры связи для генеральной совокупности равно нулю
- Дальнейшая логика проверки на статистическую значимость аналогична общей последовательности этапов проверки любой статистической гипотезы

Коэффициенты корреляции

- **Коэффициент корреляции Браве-Пирсона (r)** — это параметрический показатель, для вычисления которого сравнивают средние и стандартные отклонения результатов двух измерений
- **Коэффициент корреляции рангов Спирмена (r_s)** — это непараметрический показатель, с помощью которого пытаются выявить связь между рангами соответственных величин в двух рядах измерений

Коэффициент корреляции

- **Коэффициент корреляции** измеряет силу и направление связи между двумя переменными.
- **Обозначения:**
- Выборочный коэффициент корреляции r
- Коэффициент корреляции генеральной совокупности ρ

Коэффициент корреляции

- Коэффициент корреляции — это величина, которая может варьировать в пределах от +1 до -1 (т.е. $-1 < r < 1$)
- Коэффициенты корреляции — удобный показатель связи
- по величине коэффициентов корреляции нельзя судить о достоверности корреляционной связи между признаками
- Таблицы значений критериев можно найти в специальных руководствах

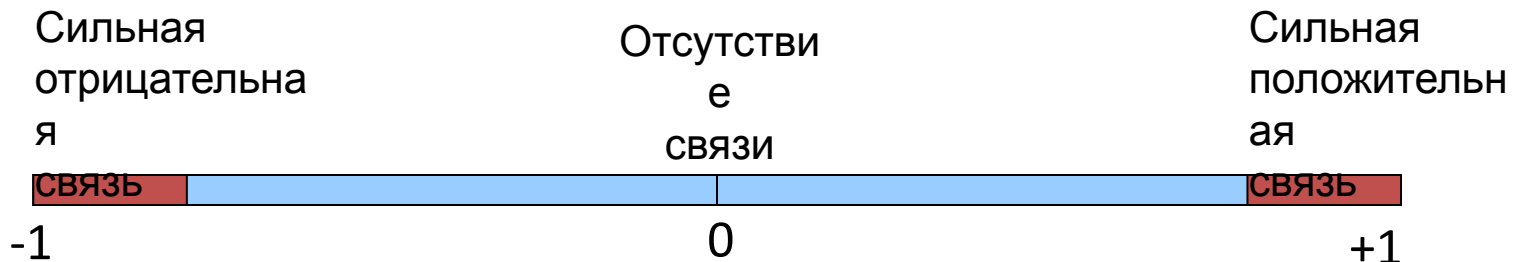
Допущения для определения коэффициента корреляции

- Метрический характер измеряемых данных (данные представлены в интервальной шкале или шкале отношений)
- Обе переменные подчиняются нормальному закону распределения
- Зависимость между переменными приблизительно линейна
- *Гомоскедастичность*, т.е. однородность дисперсий (дисперсия значений y равномерна для всех значений x). На графике значения y должны быть приблизительно равномерно распределены выше и ниже линии $y(x)$ по всей ее длине (точки на графике должны образовывать достаточно симметричную овалообразную форму без значительных выбросов)

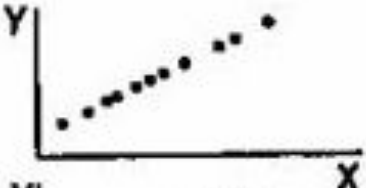


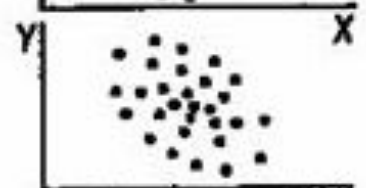
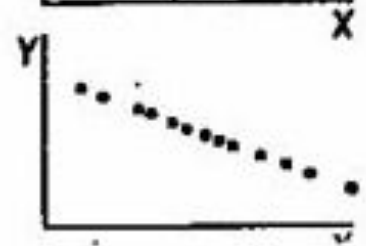
Существенные признаки нелинейности или отклонения от гомоскедастичности свидетельствуют о необходимости использовать другую меру связи и другой критерий значимости.

Значения коэффициента корреляции

- Коэффициент корреляции изменяется на отрезке от -1 до $+1$.
- Если между переменными существует сильная положительная связь, то значение r будет близко к $+1$.
- Если между переменными существует сильная отрицательная связь, то значение r будет близко к -1 .
- Когда между переменными нет линейной связи или она очень слабая, значение r будет близко к 0 .



Корреляционная связь

Интерпретация значений r_{xy}		
Величина r_{xy}	Описание линейной связи	Диаграмма рассеивания
+1,00	Строгая прямая связь	
Около +0,50	Слабая прямая связь	
0,00	Нет связи (то есть ковариация X и Y = 0)	
Около -0,50	Слабая обратная связь	
-1,00	Строгая обратная связь	

Частная классификация корреляционных связей:

- 1) высокая значимая корреляция при r , соответствующем уровню статистической значимости $\rho \leq 0,01$;
- 2) значимая корреляция при r , соответствующем уровню статистической значимости $\rho \leq 0,05$;
- 3) тенденция достоверной связи при r , соответствующем уровню статистической значимости $\rho \leq 0,10$;
- 4) незначимая корреляция при r , не достигающем уровня статистической значимости .

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ r -ПИРСОНА (r -СПИРМЕНА)

(для проверки ненаправленных альтернатив, n — объем выборки)

n	p				n	p			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428

Корреляционный анализ

1. Выбрать коррелируемые переменные (интервальная шкала, нормальное распределение)
2. Гипотеза о связи переменных
3. Корреляционная матрица
4. Коэффициенты корреляции
5. Корреляционная плеяда

Корреляционная матрица (таблица)

Workbook1* - Корреляции (база данных.sta)

Workbook1*

- Основная статистика/таб
- Correlations dialog
- Корреляции (база)

Корреляции (база данных.sta)
Marked correlations are significant at $p < ,05000$
N=15 (Casewise deletion of missing data)

Переменная	актив	общ	впеч	власть	ДЗД	и.кнф	творч	п.пом	манип	позн	извест
актив	1,00	0,56	0,73	0,61	0,82	-0,28	0,49	0,24	0,07	0,80	0,78
общ	0,56	1,00	0,60	0,38	0,39	0,09	0,31	0,40	0,16	0,46	0,60
впеч	0,73	0,60	1,00	0,53	0,68	-0,36	0,71	0,19	0,09	0,85	0,66
власть	0,61	0,38	0,53	1,00	0,38	-0,61	0,43	0,09	0,32	0,56	0,76
ДЗД	0,82	0,39	0,68	0,38	1,00	-0,36	0,43	-0,03	-0,26	0,69	0,51
и.кнф	-0,28	0,09	-0,36	-0,61	-0,36	1,00	-0,43	0,46	0,02	-0,38	-0,45
творч	0,49	0,31	0,71	0,43	0,43	-0,43	1,00	0,24	0,31	0,89	0,69
п.пом	0,24	0,40	0,19	0,09	-0,03	0,46	0,24	1,00	0,66	0,26	0,16
манип	0,07	0,16	0,09	0,32	-0,26	0,02	0,31	0,66	1,00	0,19	0,34
позн	0,80	0,46	0,85	0,56	0,69	-0,38	0,89	0,26	0,19	1,00	0,81
извест	0,78	0,60	0,66	0,76	0,51	-0,45	0,69	0,16	0,34	0,81	1,00

Корреляции (база данных.sta) | Цихончик Н.В., 2016

Коэффициенты корреляции

Корреляции

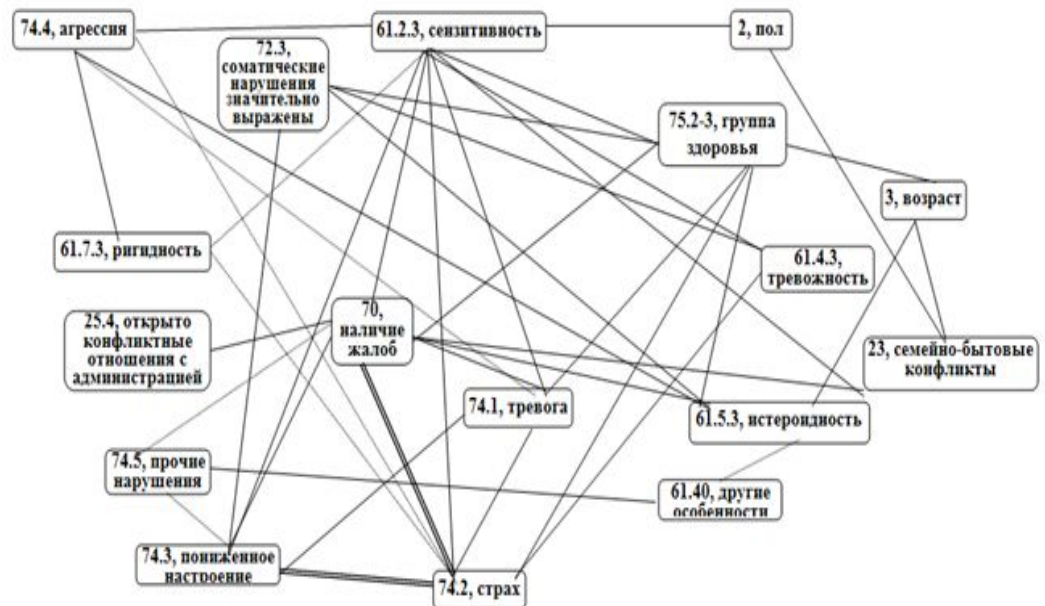
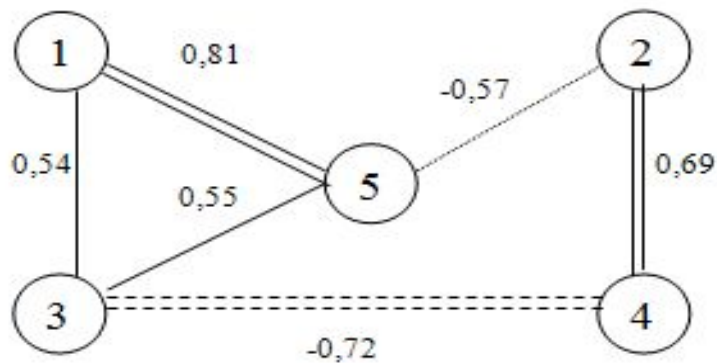
		Социнтел	Соцкомпет
Социнтел	Корреляция Пирсона	1	,782**
	Знач.(2-сторон)		,000
	N	34	34
Соцкомпет	Корреляция Пирсона	,782**	1
	Знач.(2-сторон)	,000	
	N	34	34

** . Корреляция значима на уровне 0.01 (2-сторон.).

Корреляционная плеяда

- корреляционные плеяды – это способ отображения информации о корреляциях, который помогает их структурировать, проводить объединение коррелирующих факторов
- <http://www.robotron.ru/papers/pleyadi.html>
Примеры удачных и неудачных построений плеяд

Корреляционная плеяда




Корреляционная плеяда

**Социальный
интеллект**

0,78

**Социальная
компетентность**





**2 задача занятия –
Проверка гипотез о
взаимосвязи переменных**

Excel (задача 2)

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Библиотека функций

Q3 fx =КОРРЕЛ(О3:О36;Р3:Р36)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Соц-психолог особенности личности																
2	№	ФИО	Легкость общения	Сноровка, хватка	социальная выносливость	широкая социальная компетентность	самоуважение	Понимание ситуации	Автономия	соц-проф комп-ть	социальная наблюдательность	коммун компетентность	понимание людей	эго-компетентность	Социальный интеллект	Социальная компетентность	
3	1	Безумова Алёна	17	23	13	16	15	12	13	17	12	15	10	18	80	101	0,781647
4	2	Богданова Анна	11	10	6	5	12	9	3	7	12	4	12	4	56	39	
5	3	Борисова Юлия	17	15	13	11	11	11	15	13	12	11	11	13	79	74	
6	4	Булыгина Дарья	11	13	13	12	13	12	13	11	16	12	15	15	81	75	
7	5	Волков Артур	18	17	15	12	18	19	15	11	15	14	18	21	99	94	
8	6	Гахраманова Сабина	29	24	19	15	23	20	25	13	18	23	16	25	130	120	
9	7	Громова Ольга	20	15	21	15	16	23	17	13	10	13	23	16	107	95	
10	8	Джинджиристая Виктория	21	20	18	8	18	14	18	4	8	7	16	15	99	68	
11	10	Иванова Кристина	3	3	5	7	9	13	12	9	13	5	7	13	49	50	
12	11	Ковалев Олег	14	20	21	22	19	21	21	11	17	17	21	24	113	115	
13	12	Коноплева Юлия	29	21	16	12	23	22	22	20	9	19	18	21	117	115	

SPSS

*Без имени1 [Наборданных0] - SPSS Statistics Data Editor

Файл Правка Вид Данные Преобразовать Анализ Графика Сервис Дополнения Окно Справка

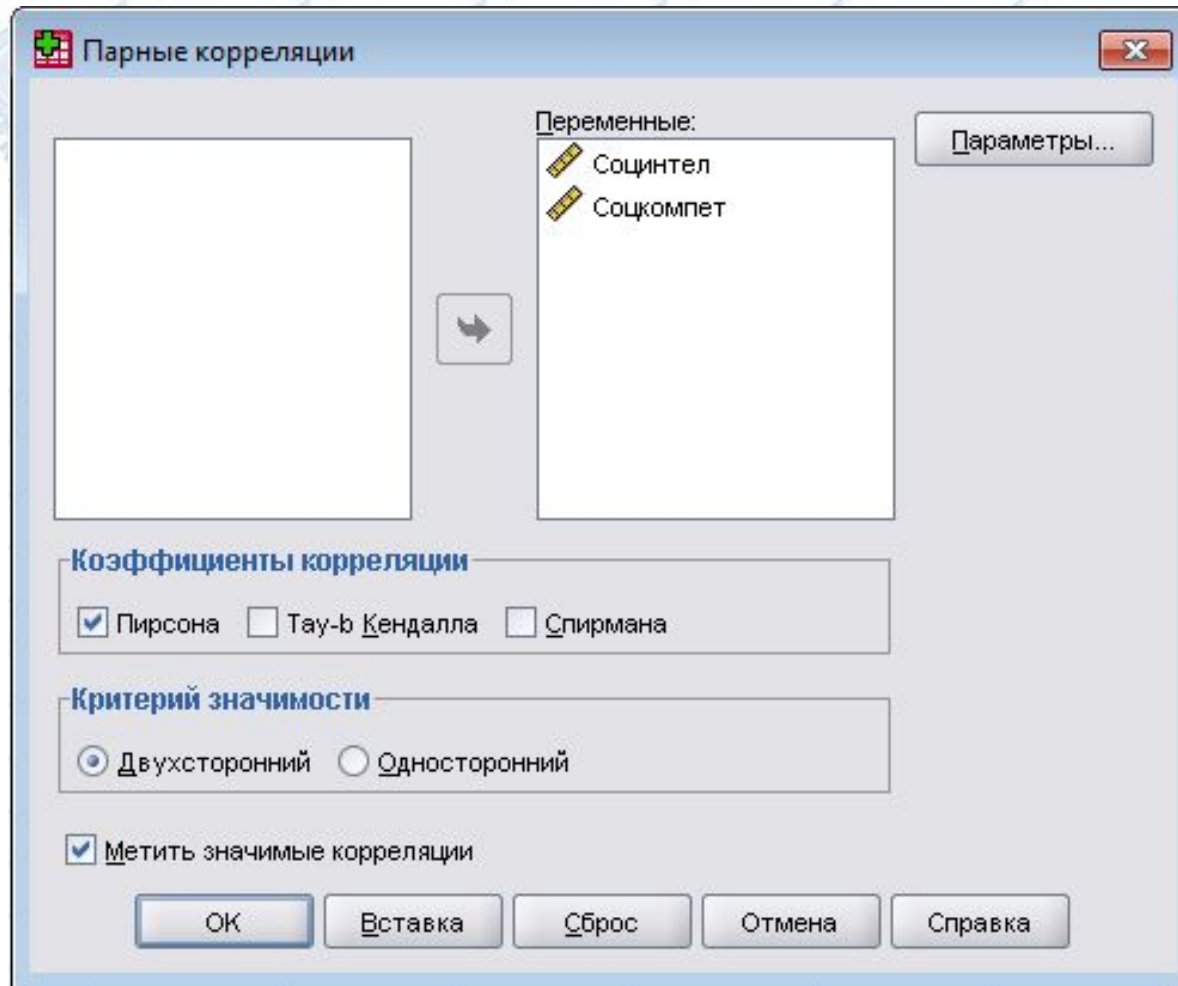
6 :

	Социнтел	Соцкомпет	п
1	80	101	
2	56	39	
3	79	74	
4	81	75	
5	99	94	
6	130	120	
7	107	95	
8	99	68	
9	49	50	
10	113	115	
11	117	115	
12	97	104	
13	119	112	
14	117	114	
15	77	77	
16	120	129	
17	102	86	
18	119	87	
19	136	122	
20	78	98	
21	91	86	
22	84	79	

Анализ

- Отчеты
- Описательные статистики
- Таблицы
- Анализ RFM
- Сравнение средних
- Общая линейная модель
- Обобщенные линейные модели
- Смешанные модели
- Корреляции**
 - r_{12} Парные...
 - $r_{12\>3}$ Частные...
 - δ Расстояния...
- Регрессия
- Логлинейный
- Нейронные сети
- Классификация
- Снижение размерности
- Шкалирование
- Непараметрические критерии
- Прогнозирование
- Выживаемость
- Множественные ответы
- Анализ пропущенных значений...
- Множественная импутация
- Сложные выборки
- Контроль качества
- ROC кривые...

SPSS



Задания

1. Задача 6. Построить корреляционную плеяду для групп 1 и 2.
2. Задача 8. Построить корреляционную плеяду для групп 1 и 2 по 5 стратегиям поведения в конфликте.
3. Задача 10. Установить значимые корреляции успеваемости с показателями теста.

Задание к следующему занятию

1. Кластерный анализ: понятие и назначение процедуры.
2. Виды кластерного анализа.