

Выбор критерия для статистического анализа планируемого исследования

доцент кафедры
«Эпидемиологии, доказательной медицины и
биостатистики» КМУ ВШОЗ,
PhD., Сыдықова Б.Қ
b.sydykova@ksph.kz

Выбор критерия

Переменная 1	Переменная 2	Анализ
Количественная	Качественная	Алгоритм анализа количественных переменных
Количественная	Количественная	Корреляционный анализ
Качественная	Качественная	Алгоритм анализа качественных переменных
Изучаемая (зависимая) переменная	Переменные, влияние которых на зависимую мы изучаем (независимые переменные)	Регрессионный анализ

Анализ количественных данных

Количественная переменная

- Проверить распределение.
- Описать переменную (M, Me, SD, IQR)
- ✓ Нормальное распределение – M (SD)
- ✓ Распределение отличное от нормального –
M (Me, Q1, Q3) или M (Me, IQR)
- Выбрать критерий
- Провести расчёт
- Интерпретировать
- Сделать вывод

Подчиняются ли данные закону нормального распределения?

Да

Нет

Являются ли группы независимыми, связанными или проводится сравнение с заданным значением?

Независимые

Связанные

Сравнение с
заданным
значением

Независимые

Связанные

Сравнение с
заданным
значением

Сколько групп сравниваем?

2

3 или больше

2

3 или больше

1

2

3 или больше

2

3 или больше

1

t-критерий Стьюдента для
независимых выборок

Однофакторный
дисперсионный анализ

t-критерий Стьюдента для
парных выборок

Дисперсионный анализ для
повторных наблюдений

t-критерий Стьюдента для
одной выборки

Критерий Манна-Уитни

Критерий Краскела-Уоллиса

Критерий Вилкоксона

Критерий Фридмана

Критерий Вилкоксона

Нормальное распределение

- Q-Q диаграмма

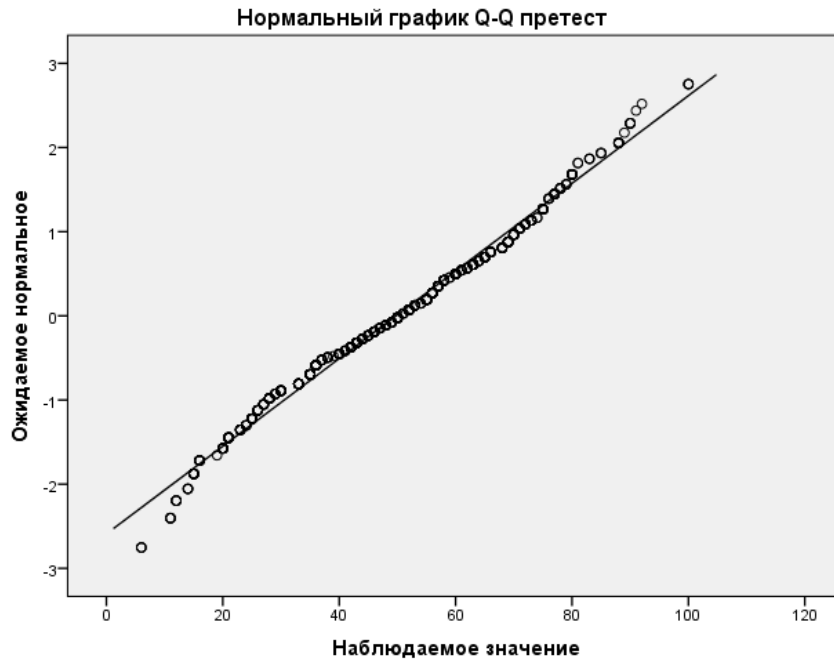
- Гистограмма

- Коэффициент асимметрии –

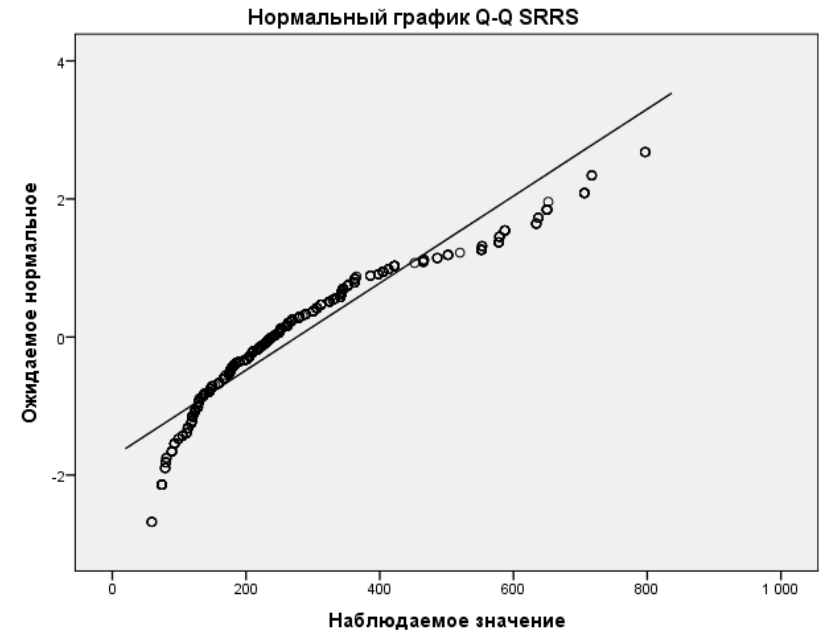
Значение от 0 до 1 – нормальное распределение,
более 1 – отличается от нормального

- Равенство Средней и Медианы

Q-Q диаграмма

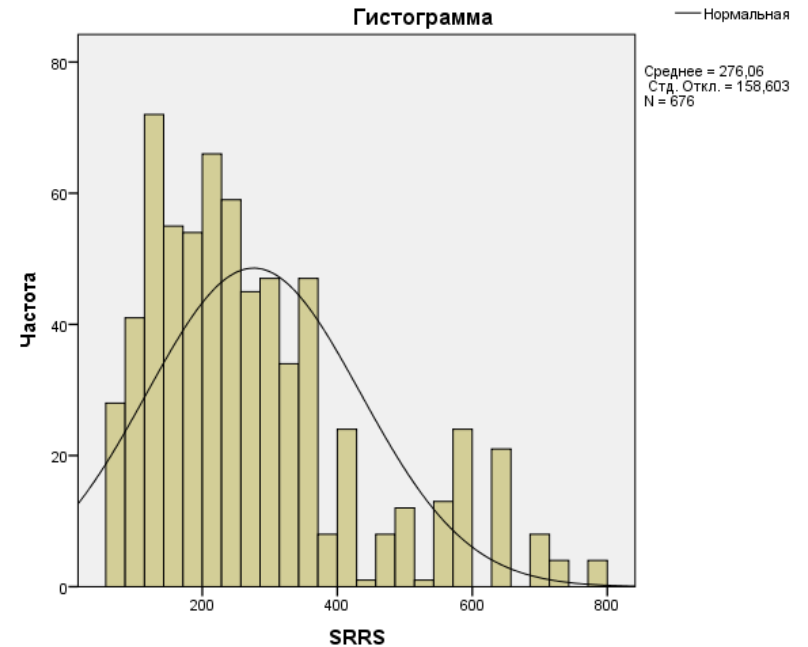
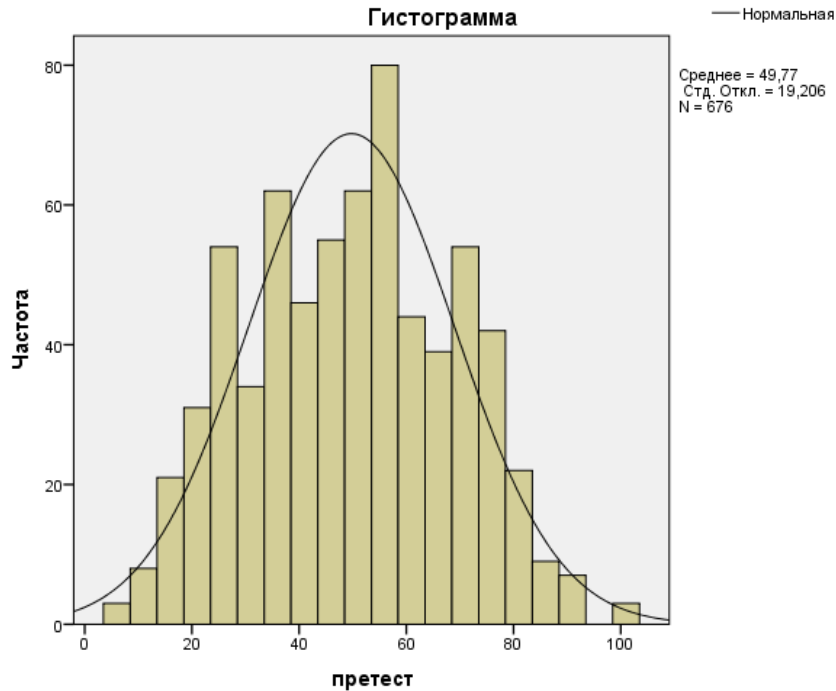


- Нормальное распределение



- Отличающееся от нормального распределение

Гистограмма



- Нормальное распределение

- Отличающееся от нормального распределение

Коэффициент асимметрии М и М_е

- SRRS

Претест

Среднее	49,77	,739
95% Доверительный интервал для среднего	Нижняя граница	48,32
	Верхняя граница	51,22
Среднее по выборке, усеченной на 5%	49,75	
Медиана	50,00	
Дисперсия	368,887	
Среднекв.отклонение	19,206	
Минимум	6	
Максимум	100	
Диапазон	94	
Межквартильный диапазон	30	
Асимметрия	,002	,094
Эксцесс	-,695	,188

Описательные статистики

	Статистика	Стандартная ошибка
Среднее	276,06	6,100
95% Доверительный интервал для среднего	Нижняя граница	264,08
	Верхняя граница	288,04
Среднее по выборке, усеченной на 5%	264,40	
Медиана	238,00	
Дисперсия	25154,851	
Среднекв.отклонение	158,603	
Минимум	59	
Максимум	797	
Диапазон	738	
Межквартильный диапазон	186	
Асимметрия	1,115	,094
Эксцесс	,713	,188

- Нормальное распределение

- Отличающееся от нормального распределение

Подчиняются ли данные закону нормального распределения?

Да

Нет

Являются ли группы независимыми, связанными или проводится сравнение с заданным значением?

Независимые

Связанные

Сравнение с
заданным
значением

Независимые

Связанные

Сравнение с
заданным
значением

Сколько групп сравниваем?

2

3 или больше

2

3 или больше

1

2

3 или больше

2

3 или больше

1

t-критерий Стьюдента для
независимых выборок

Однофакторный
дисперсионный анализ

t-критерий Стьюдента для
парных выборок

Дисперсионный анализ для
повторных наблюдений

t-критерий Стьюдента для
одной выборки

Критерий Манна-Уитни

Критерий Краскела-Уоллиса

Критерий Вилкоксона

Критерий Фридмана

Критерий Вилкоксона

Кратковременная память и стресс у студентов медицинского университета

группа_иссл_контр

	Частота	Процент
основная группа	367	54,3
контрольная группа	309	45,7
Итого	676	100,0

семейное_положение

	Частота	Процент	Всего
холост	613	90,7	
замужем	63	9,3	
Итого	676	100,0	

пол

	Частота	Процент
жен	356	52,7
муж	320	47,3
Итого	676	100,0

курс

	Частота	Процент
1 курс	239	35,4
3 курс	203	30,0
5 курс	234	34,6
Итого	676	100,0

проживание

	Частота	Процент
собственное жилье	404	59,8
аренда	135	20,0
живет в общежитии	137	20,3
Итого	676	100,0

закончил_школу

	Частота	Процент
город	551	81,5
село	125	18,5
Итого	676	100,0

дети

	Частота	Процент
нет	641	94,8
есть	35	5,2
Итого	676	100,0

Кратковременная память и стресс у студентов медицинского университета

- **SRRS** – количественная переменная, обозначающая уровень стресса у студентов, измеряемая анкетой Хомса-Рэге (балл)
- **Пре-тест, пост-тест** – количественные переменные, обозначающие уровень кратковременной памяти, измеренные при помощи программы N-back (% правильных ответов, min = 0, max=100)
- **GPA-before, GPA-after** – количественная переменная до и после окончания тренировок памяти студентов в основной группе. Оценивался как в основной, так и в контрольной группе.

План

- SRRS, пре-тест, пост-тест
- Группа исследования, пол, курс

SRRS

Курс

- 1 курс: $M=278,8$ ($Me=249,0$; $IQR=199,0$)
- 3 курс: $M=260,7$ ($Me=228,0$; $IQR=196,0$)
- 5 курс: $M=272,1$ ($Me=243,0$; $IQR=197,0$)

Статистически значима ли разница? Какой критерий необходимо использовать?

SRRS

Курс

- 1 курс: $M=278,8$ ($Me=249,0$; $IQR=199,0$)
- 3 курс: $M=260,7$ ($Me=228,0$; $IQR=196,0$)
- 5 курс: $M=272,1$ ($Me=243,0$; $IQR=197,0$)

Критерий Краскела-Уоллиса

$H=2,091$; $p=0,351$

SRRS

Пол

- Жен: $M=260,8$ ($Me=219,5$; $IQR=196,0$)
- Муж: $M=293,0$ ($Me=267,0$; $IQR=196,0$)

Статистически значима ли разница? Какой критерий необходимо использовать?

SRRS

Пол

- Жен: $M=260,8$ ($Me=219,5$; $IQR=196,0$)
- Муж: $M=293,0$ ($Me=267,0$; $IQR=196,0$)

Критерий Манна-Уитни

$U=49230,0$; $p=0,002$

Пре-тест

Пол

- Жен: $M=49,54$ ($SD=18,8$)
- Муж: $M=50,02$ ($SD=19,7$)

Статистически значима ли разница? Какой критерий необходимо использовать?

Пре-тест

Пол

- Жен: $M=49,54$ ($SD=18,8$)
- Муж: $M=50,02$ ($SD=19,7$)

t-критерий Стьюдента для независимых выборок = $-0,318$; $df=674$; $p=0,751$

Пре-тест

Курс

- 1 курс: $M=51,56$ ($SD=19,1$)
- 3 курс: $M=48,67$ ($SD=19,8$)
- 5 курс: $M=48,89$ ($SD=18,7$)

Статистически значима ли разница? Какой критерий необходимо использовать?

Пре-тест

Курс

- 1 курс: $M=51,56$ ($SD=19,1$)
- 3 курс: $M=48,67$ ($SD=19,8$)
- 5 курс: $M=48,89$ ($SD=18,7$)

Однофакторный дисперсионный анализ

$F=1,622$; $p=0,198$

Пост-тест

Группа исследования

- Основная: $M=59,23$ ($SD=19,1$)
- Контрольная: $M=53,77$ ($SD=19,4$)

Статистически значима ли разница? Какой критерий необходимо использовать?

Пост-тест

Группа исследования

- Основная: $M=59,23$ ($SD=19,1$)
 - Контрольная: $M=53,77$ ($SD=19,4$)
 - t-критерий Стьюдента для независимых выборок = $3,673$; $df=674$; $p<0,001$
-

Пост-тест

Группа исследования

- Основная: $M=59,23$ ($SD=19,1$)
 - Контрольная: $M=53,77$ ($SD=19,4$)

 - t-критерий Стьюдента для независимых выборок = $3,673$; $df=674$; $p<0,001$
-

Пре-тест / Пост-тест

Группа исследования *(по отдельности)*

- Основная: $M=46,98$ ($SD=18,9$) /
 $M=59,23$ ($SD=19,1$)
- Контрольная: $M=53,08$ ($SD=19,0$) /
 $M=53,77$ ($SD=19,4$)
- Статистически значима ли разница? Какой критерий необходимо использовать?

Пре-тест / Пост-тест

Группа исследования

- Основная: $M=46,98$ ($SD=18,9$) /
 $M=59,23$ ($SD=19,1$)
- Контрольная: $M=53,08$ ($SD=19,0$) /
 $M=53,77$ ($SD=19,4$)
- t-критерий Стьюдента для парных выборок
= 31,943; $df=366$; $p<0,001$ (основная)
- t-критерий Стьюдента для парных выборок
= 1,846; $df=308$; $p=0,066$ (контрольная)

Подчиняются ли данные закону нормального распределения?

Да

Нет

Являются ли группы независимыми, связанными или проводится сравнение с заданным значением?

Независимые

Связанные

Сравнение с
заданным
значением

Независимые

Связанные

Сравнение с
заданным
значением

Сколько групп сравниваем?

2

3 или больше

2

3 или больше

1

2

3 или больше

2

3 или больше

1

t-критерий Стьюдента для
независимых выборок

Однофакторный
дисперсионный анализ

t-критерий Стьюдента для
парных выборок

Дисперсионный анализ для
повторных наблюдений

t-критерий Стьюдента для
одной выборки

Критерий Манна-Уитни

Критерий Краскела-Уоллиса

Критерий Вилкоксона

Критерий Фридмана

Критерий Вилкоксона

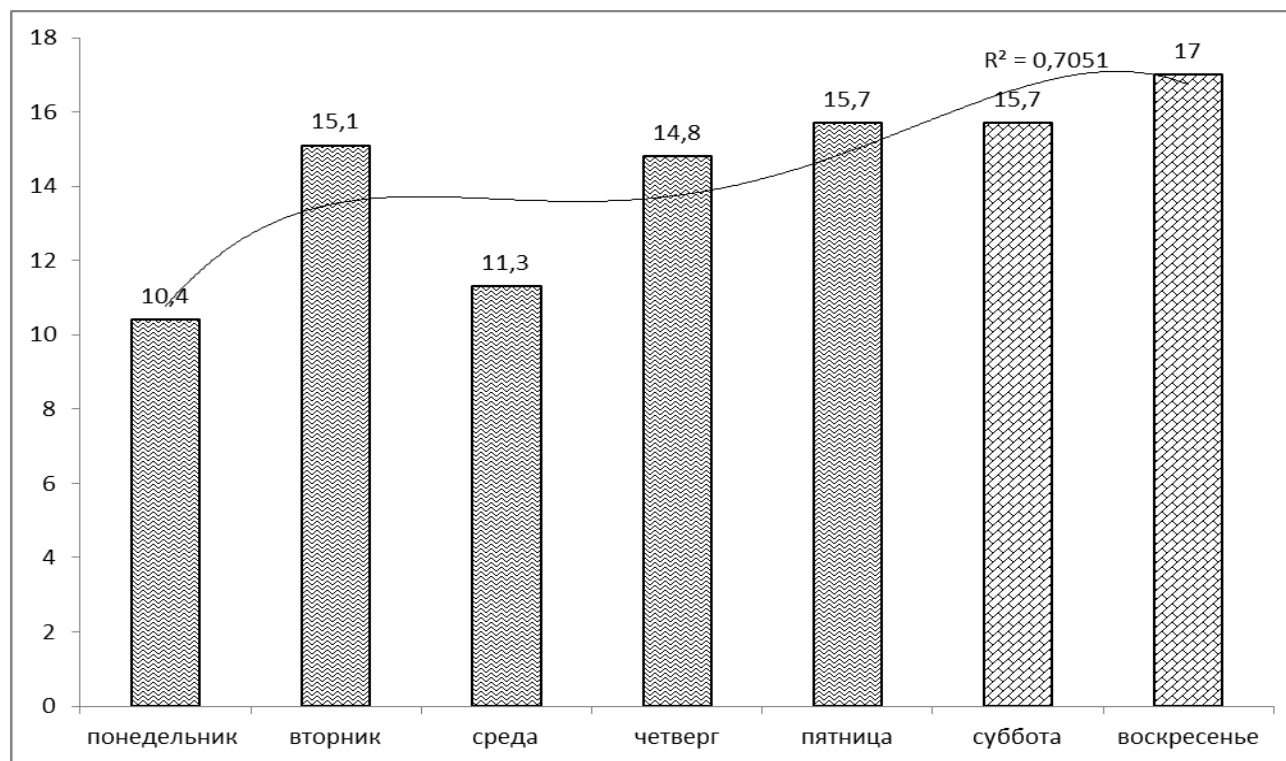
Анализ качественных данных

Качественные данные

- Дихотомические (0,1)
- Номинальные (А,В,С)
- Порядковые (I, II, III)

Хи-квадрат Пирсона

На столбиковой диаграмме представлена средняя сезонность летальных случаев в течение недели (%) от ДТП.

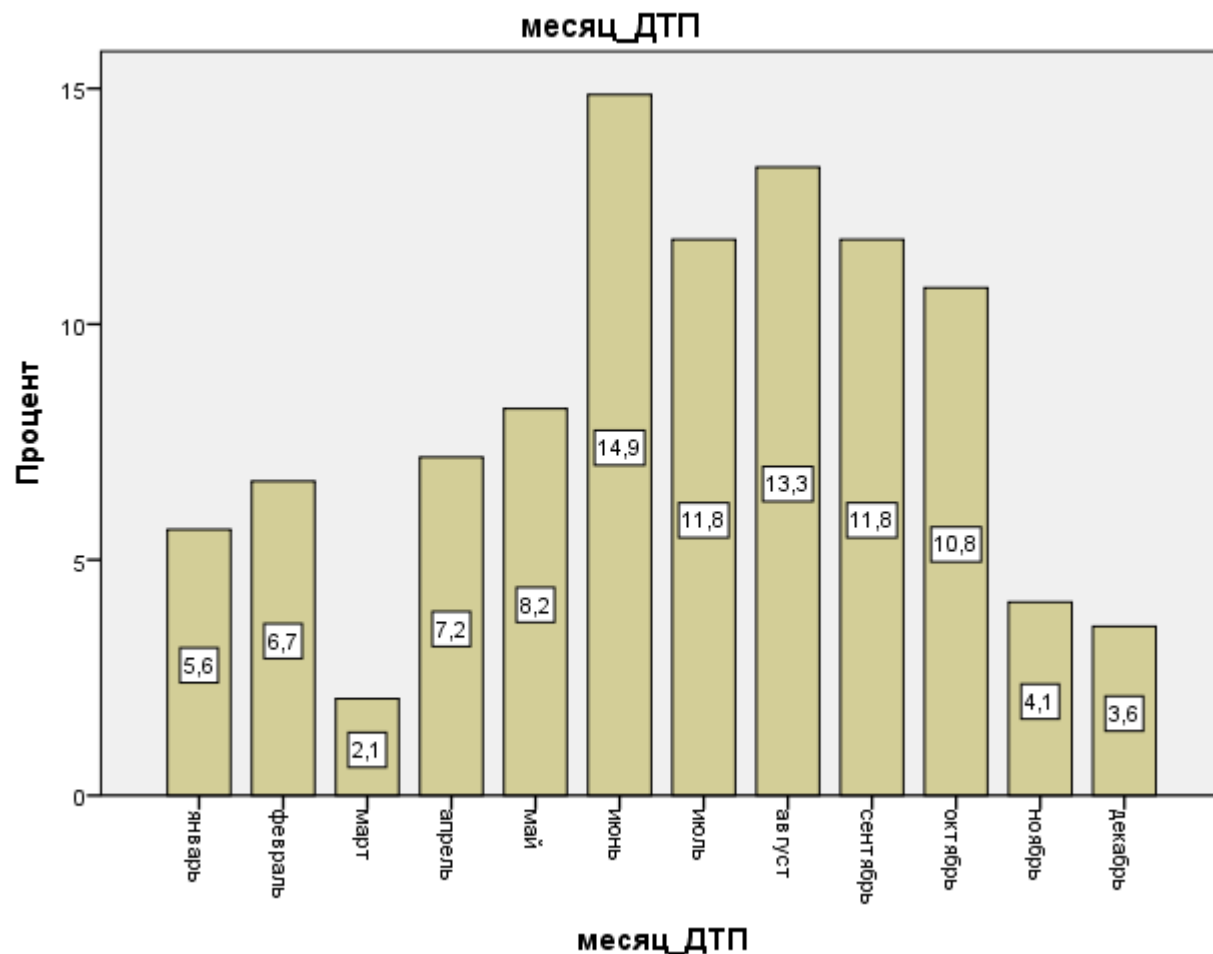


Статистики критерия

	день недели ДТП
Хи-квадрат	13,456 ^а
ст.св.	6
Асимпт. знч.	,036

а. Частоты, меньшие 5, ожидалось в 0 ячейках (0,0%). Минимальная ожидаемая частота равна 27,9.

Хи-квадрат Пирсона



Статистики критерия

	месяц_ДТП
Хи-квадрат	44,200 ^a
ст.св.	11
Асимпт. знч.	,000

a. Частоты, меньшие 5, ожидалось в 0 ячейках (0,0%).
Минимальная ожидаемая частота равна 16,3.

Что показывает критерий χ^2

- Есть ли взаимосвязь между качественными переменными
- Не показывает где эти различия в случае многопольных таблиц
- Не показывает силу взаимосвязи

Что показывают силу взаимосвязи?

- Phi (От 0 до 1. Только для 2x2 таблиц)
- Коэффициент сопряженности (Contingency coefficient) Редко достигает максимума (1)
- Cramer's V (от 0 до 1)
- Отношение шансов (Odds ratio)

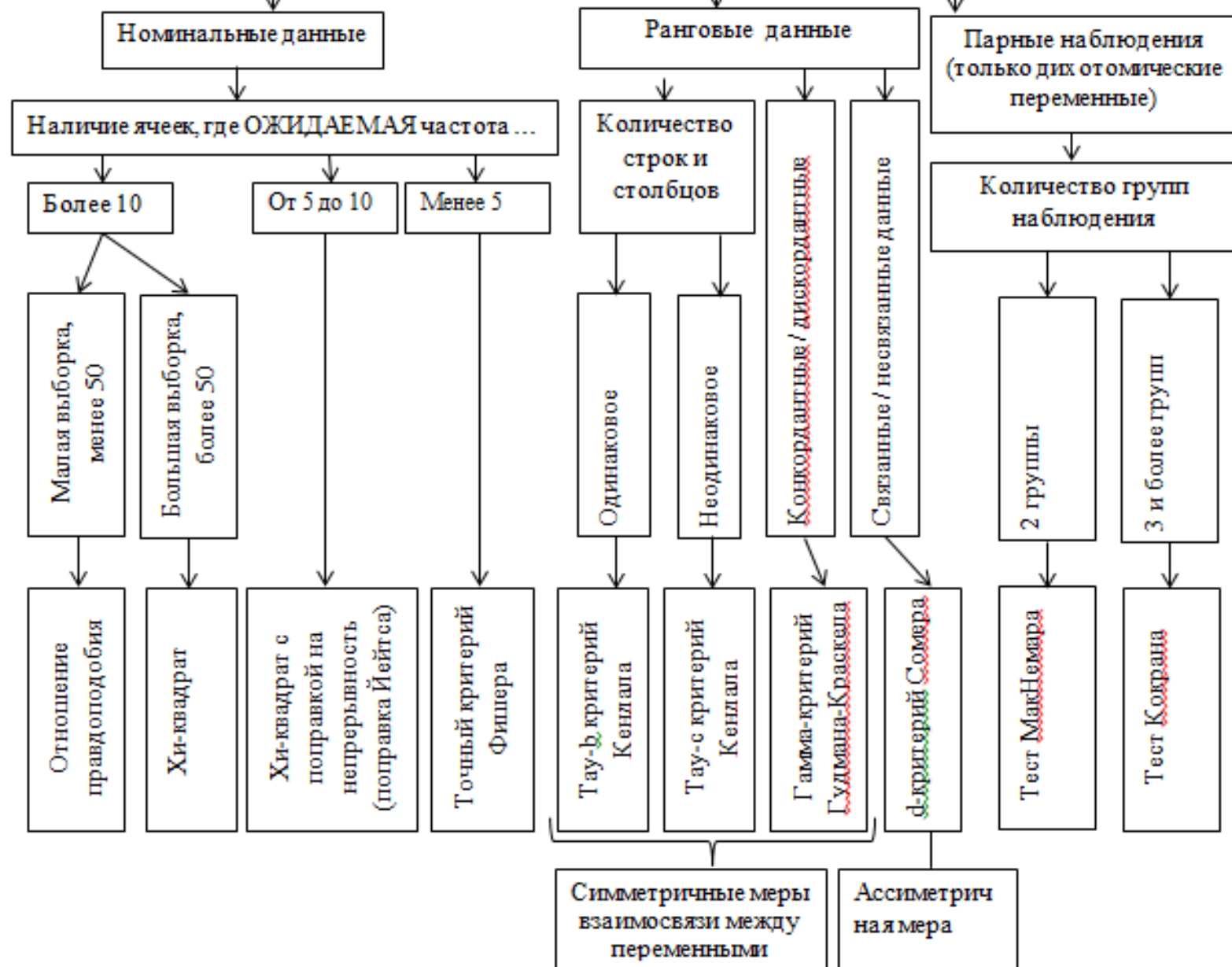
Критерии для ранговых переменных

- Kendall's tau-b
- Kendall's tau-c
- Somer's d
- Gamma
- Значения для всех критериев варьирует от -1 до 1.
- Интерпретация как для корреляционного анализа

Аналоги для парных наблюдений

- McNemar test для двух групп
- Analyze->Non-parametric tests->2 related samples.
Выбрать McNemar
- Cochran's Q-test для 3 и более
- Analyze->Non-parametric tests->K related samples.
Выбрать Cochran's Q
- Оба критерия – только для дихотомических переменных (0:1)

Качественные данные



Корреляционный анализ

Корреляционный анализ

Показывает только меру ассоциации !

Есть или нет (значение p) + силу
(коэффициент) + направление (знак)

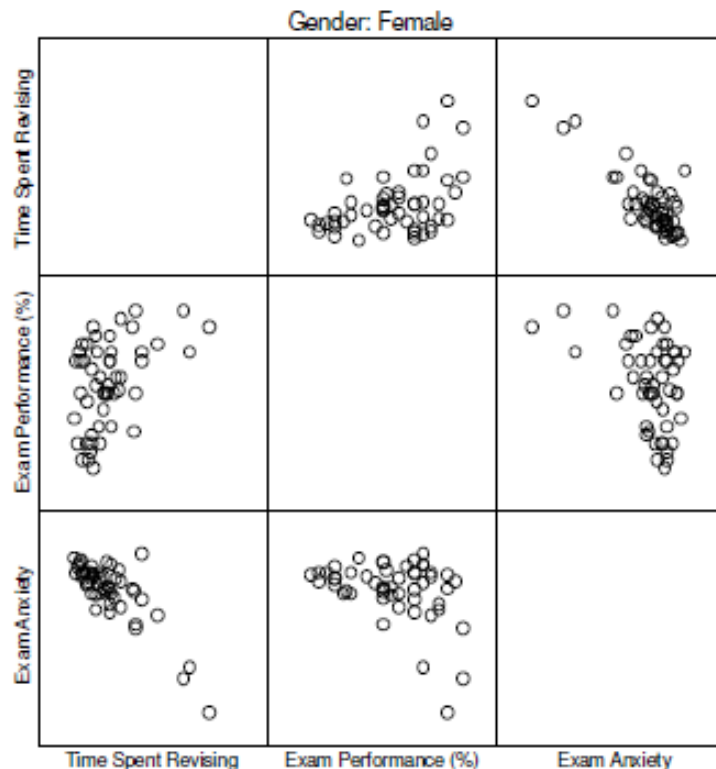
Корреляционная	$\neq$	причинно-
связь		
следственная		
		связь

Корреляционный анализ

Две количественные переменные

- Обе нормально распределены – коэффициент корреляции Пирсона
- Одна или обе – отличается от нормального – коэффициент корреляции Спирмена
- «+» - прямо пропорциональная связь
- «-» – обратно пропорциональная связь

Корреляционная матрица

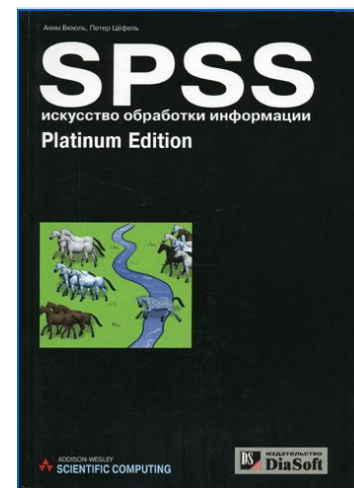


- Коэффициент корреляции всегда между -1 и 1.
- 0 – нет связи
- 1 – идеальная прямопропорциональная зависимость
- -1 - идеальная обратнопропорциональная зависимость
- 0.7 – сильная связь ?
- 0.5 – средней силы ?
- 0.3 – слабая связь ?

Ахим Бююль, Петер Цефель

SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей.

SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows



Значение коэффициента корреляции r	Интерпретация
$0 < r \leq 0,2$	Очень слабая корреляция
$0,2 < r \leq 0,5$	Слабая корреляция
$0,5 < r \leq 0,7$	Средняя корреляция
$0,7 < r \leq 0,9$	Сильная корреляция
$0,9 < r \leq 1$	Очень сильная корреляция

Регрессионный анализ

Регрессионный анализ

- Переменные (факторы), влияние которых мы изучаем, могут быть любыми – независимые переменные.
- Переменная (исход), влияние на которую других переменных (факторов) – зависимая переменная.

Изучаемая (зависимая) переменная	Название регрессионного анализа
Количественная	Линейный регрессия
Порядковая	Регрессионный анализ в порядковой переменной отклика
Номинальная	Мультиномиальная логистический регрессия
Дихотомическая	Бинарная логистическая регрессия

Выбор критерия

Переменная 1	Переменная 2	Анализ
Количественная	Качественная	Алгоритм анализа количественных переменных
Количественная	Количественная	Корреляционный анализ
Качественная	Качественная	Алгоритм анализа качественных переменных
Изучаемая (зависимая) переменная	Переменные, влияние которых на зависимую мы изучаем (независимые переменные)	Регрессионный анализ

Спасибо за внимание!
