

## Тема 5: «СРАВНЕНИЕ СОВОКУПНОСТЕЙ ПО НЕПАРАМЕТРИЧЕСКИМ КРИТЕРИЯМ»

Параметрические методы статистики используют параметры нормального распределения – среднее арифметическое и дисперсию. Следовательно, чтобы применять эти методы нужно убедиться в нормальности имеющегося эмпирического распределения. Если эмпирическое распределение не соответствует нормальному, то среднее арифметическое и дисперсия (или стандартное отклонение,  $\sigma$ ) дадут о нем неверное представление. Неверными окажутся и статистические критерии, рассчитанные на основе этих параметров. Иногда выборка бывает очень мала, что не позволяет обнаружить нормальность распределения, поэтому воспользоваться параметрическими критериями тоже не представляется возможным. В таких случаях применяют **непараметрические критерии**.

При использовании непараметрических методов нет необходимости доказывать нормальность эмпирического распределения. Достаточно оценить различие двух (или более) эмпирических распределений. Многие непараметрические критерии основаны на ранжировании переменной. Фактические значения признака упорядочиваются, и далее работа выполняется уже с рангами. При этом сохраняется большая часть информации об эмпирическом распределении, но сам вид распределения значения уже не имеет.

Наиболее часто используются следующие непараметрические критерии (обычно они носят имена предложивших их математиков):

- I. Критерий Манна-Уитни (U)
- II. Критерий Уилкоксона (W)
- III. Критерий Пирсона  $\chi^2$  (читается «хи-квадрат»)
- IV. Критерий МакНемара

### I. КРИТЕРИЙ МАННА-УИТНИ (U)

Критерий Манна-Уитни используется для сравнения двух *несвязанных* выборок малого объема. По сути, он является непараметрическим аналогом t-критерия Стьюдента для несвязанных групп. Фактические значения признака ранжируются, далее расчеты производят с рангами, так что характер эмпирического распределения уже неважен.

#### Задача 1

В эксперименте с параллельным отрицательным контролем изучено новое лекарственное средство, предположительно обладающее диуретическим эффектом. В контрольной группе принимавших плацебо 3 человека, в основной группе получавших препарат 4 человека. Величина суточного диуреза у каждого испытуемого представлена в таблице 1. Необходимо оценить статистическую значимость различия диуреза в группах сравнения.

Таблица 1.

#### Суточный диурез у испытуемых в контрольной и основной группах

Суточный диурез (мл) в контрольной группе (получали плацебо)	Суточный диурез (мл) в основной группе (получали препарат)
1000	1400
1380	1600
1200	1180
	1220

Из соображений наглядности можно рассчитать средние арифметические для этих двух групп наблюдений. Средний суточный диурез в группе лиц, получавших плацебо,

составил 1193 мл ( $M_1$ ), а в группе получавших препарат 1350 мл ( $M_2$ ). Однако применять t-критерия Стьюдента для несвязанных групп нельзя: сравниваемые группы слишком малы, чтобы можно было судить о нормальности распределения.

### Решение.

#### I. Выдвинуть нулевую и альтернативную гипотезы

- Нулевая гипотеза ( $H_0$ ): группы не имеют статистически значимых различий (т.е. препарат не обладает диуретическим эффектом).  $H_0: M_1=M_2$
- Альтернативная гипотеза ( $H_1$ ): группы наблюдений различаются (т.е. препарат обладает диуретическим эффектом)  $H_1: M_1 \neq M_2$

#### II. Рассчитать эмпирическое значение критерия Манна-Уитни

1. Данные обеих групп мысленно объединить в одну группу и ранжировать, причем обязательно от меньшего значения к большему. Ранг «1» присвоить наименьшему значению, ранг «2» – следующему по возрастанию и т.д. Наибольший ранг присваивают самому большому среди значений в обеих группах. Если значения совпадают, им присвоить один и тот же средний ранг (например, если два значения поделили 3-е и 4-е места, обоим присвоить ранг «3,5»). Результаты ранжирования представлены в таблице 2.

Таблица 2

#### Величина суточного диуреза и ранги в контрольной и основной группах

1. Контрольная группа (получали плацебо)		2. Основная группа (получали препарат)	
Суточный диурез (мл)	Ранги	Суточный диурез (мл)	Ранги
1000	1	1400	6
1380	5	1600	7
1200	3	1180	2
		1220	4

2. Подсчитать сумму рангов отдельно для каждой группы. В нашем примере:
  - контрольная группа  $1 + 5 + 3 = 9$
  - основная группа  $6 + 7 + 2 + 4 = 19$
3. Определить группу с наибольшей суммой рангов. В нашем примере это вторая группа, состоящая из 4 единиц ( $n_x$ ). Для нее сумма рангов равна 19 ( $T_x$ ).
4. Рассчитать эмпирическое значение U-критерия Манна-Уитни по формуле:

$$U = n_1 * n_2 + \frac{n_x * (n_x + 1)}{2} - T_x$$

где

$n_1$  – число наблюдений в первой группе

$n_2$  – число наблюдений во второй группе

$n_x$  – число наблюдений в группе с наибольшей суммой рангов

$T_x$  – наибольшая сумма рангов

$$U_{\text{эмп}} = 3 * 4 + \frac{4 * (4 + 1)}{2} - 19 = 12 + 10 - 19 = 3$$

#### III. Оценить эмпирическое значение критерия Манна-Уитни и сделать вывод

1. Определить критическое значение критерия Манна-Уитни по таблице (таблица 3). Для этого необходимо:
  - определить интересующий исследователя уровень значимости ( $p=0,05$  или  $p=0,01$ ) и выбрать соответствующую часть таблицы (выбираем  $p=0,05$ );
  - найти критическое значение для групп данной численности (в нашем примере  $U_{\text{кр}} = 0$ )

Критические значения критерия Манна-Уитни ( $U$ )<sup>1</sup>

Критические значения для 5% ошибки ( $p=0,05$ )																			
$n_1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n_2$																			
3	...	0																	
4	...	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								
13	2	6	10	15	19	24	28	33	37	42	47	51							
14	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
Критические значения для 1% ошибки ( $p=0,01$ )																			
5	...	...	0	1															
6	...	...	1	2	3														
7	...	0	1	3	4	6													
8	...	0	2	4	6	7	9												
9	...	1	3	5	7	9	11	14											
10	...	1	3	6	8	11	13	16	19										
11	...	1	4	7	9	12	15	18	22	25									
12	...	2	5	8	11	14	17	21	24	28	31								
13	0	2	5	9	12	16	20	23	27	31	35	39							
14	0	2	6	10	13	17	22	26	30	34	38	43	47						
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56					
16	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66				
17	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77			
18	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88		
19	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	
20	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

<sup>1</sup> Таблицы критических значений для групп, численность которых больше 20 единиц, вы сможете найти в учебниках и учебных пособиях по математической статистике, а также в глобальной сети Интернет.

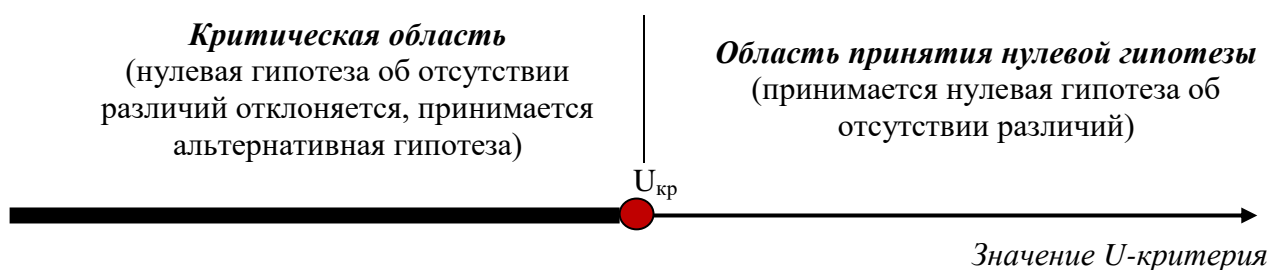
2. Сравнить эмпирическое значение U-критерия Манна-Уитни ( $U_{\text{эмп}}$ ) с критическим значением ( $U_{\text{кр}}$ )

*Особенность этого критерия состоит в том, что критическая область располагается не справа, а слева от критического значения (рисунок 1)*

- Если эмпирическое значение U-критерия ( $U_{\text{эмп}}$ ) меньше критического ( $U_{\text{кр}}$ ) или равно ему, то признается существование статистически значимых различий между уровнем признака в сравниваемых группах (т.е. принимается альтернативная гипотеза,  $H_1$ )
- Если же эмпирическое значение U-критерия ( $U_{\text{эмп}}$ ) больше критического ( $U_{\text{кр}}$ ), то принимается нулевая гипотеза об отсутствии статистически значимых различий. Значимость различий будет тем выше, чем меньше эмпирическое значение U-критерия.

Рисунок 1

### Критическое значение U-критерия Манна-Уитни и формируемые им области



- В нашем примере численность сравниваемых групп настолько мала, что мы могли бы достаточно уверенно (хотя бы с 95%-й надежностью,  $p=0,05$ ) отклонить нулевую гипотезу только получив эмпирическое значение критерия, равное нулю. В нашем примере  $U_{\text{эмп}} = 3$ , что соответствует области принятия нулевой гипотезы.

**Вывод:** Для отклонения нулевой гипотезы о равенстве суточного диуреза у пациентов сравниваемых групп достаточных аргументов не получено. Диуретический эффект у изучаемого лекарственного препарата не выявлен (статистически значимых различий суточного диуреза не обнаружено,  $p > 0,05$ ).

## II. КРИТЕРИЙ УИЛКОКСОНА (W)

Критерий Уилкоксона (или иногда переводят «критерий Вилкоксона») используется для сравнения *связанных* выборок. Он аналогичен парному t-критерию Стьюдента для связанных групп. В соответствии с методикой расчета этого критерия сначала необходимо определить величину изменения изучаемого количественного признака по каждой паре наблюдений, а затем эти изменения ранжировать.

### Задача 2

Изучается новое лекарственное средство, предположительно обладающее диуретическим эффектом. У 6 испытуемых измерили суточный диурез до приема препарата, а затем – после приема препарата (таблица 4). Как видим, у 5 из 6 испытуемых диурез увеличился. Но является ли это надежным доказательством того, что препарат обладает диуретическим эффектом?

Для ответа на этот вопрос оценим статистическую значимость изменения суточного диуреза в изучаемой группе испытуемых.

**Суточный диурез у испытуемых до и после приема препарата**

Испытуемый	Суточный диурез до приема препарата (мл)	Суточный диурез после приема препарата (мл)
A	1490	1600
B	1300	1850
C	1250	1250
D	1400	1300
E	1410	1500
F	1350	1400
G	1000	1010
H	1200	1150

Из соображений наглядности можно рассчитать средний суточный диурез у пациентов до и после приема препарата. Он составляет соответственно 1325 мл ( $M_1$ ) и 1443 мл ( $M_2$ ). Применять t-критерия Стьюдента для несвязанных групп нельзя: сравниваемые группы наблюдений представляют собой характеристики одних и тех же испытуемых, только в двух разных состояниях (т.е. сравниваемые группы являются связанными). Однако применять парный t-критерий Стьюдента для связанных групп нельзя тоже: выборка слишком мала, что не позволит убедительно продемонстрировать нормальность распределения.

**Решение.****I. Выдвинуть нулевую и альтернативную гипотезы**

- *Нулевая гипотеза ( $H_0$ ):* различия сравниваемых групп наблюдений статистически незначимы (т.е. препарат не обладает диуретическим эффектом);  $H_0: M_1=M_2$
- *Альтернативная гипотеза ( $H_1$ ):* различия сравниваемых групп наблюдений статистически значимы (т.е. препарат является диуретиком);  $H_1: M_1 \neq M_2$

**II. Рассчитать эмпирическое значение критерия Уилкоксона (W)**

1. Рассчитаем изменения признака в каждой паре наблюдений (столбец 4 таблицы 5). Пары наблюдений, где не произошло изменений признака, в дальнейшем анализе не учитываются.

Таблица 5

**Величина изменений суточного диуреза и соответствующие им ранги**

Испытуемый	Суточный диурез до приема препарата (мл)	Суточный диурез после приема препарата (мл)	Величина изменения диуреза (мл)	Ранг изменения диуреза	Ранг изменения диуреза со знаком
1	2	3	4	5	6
A	1490	1600	+110	6	+6
B	1300	1850	+550	7	+7
C	1250	1250	0	–	–
D	1400	1300	–100	5	–5
E	1410	1500	+90	4	+4
F	1350	1400	+50	2,5	+2,5
G	1000	1010	+10	1	+1
H	1200	1150	–50	2,5	–2,5

2. Ранжируем абсолютные величины изменения признака (столбец 5 таблицы 5), соблюдая следующие правила:

- из ранжирования исключаются единицы, у которых не произошло никаких изменений (в нашем примере испытуемый С),
- направление изменения признака не учитывается (то есть пока нам неважно, увеличился или уменьшился диурез у испытуемых),
- ранжирование выполняется *от меньшего к большему*,
- одинаковым значениям присваивают одинаковый средний ранг.

В нашем примере у испытуемых F и H диурез изменился на одинаковую величину (50 мл). Следовательно, и ранги должны быть одинаковыми. При ранжировании этих величин должны быть использованы ранги 2 и 3, значит каждой из них присвоим одинаковый средний ранг 2,5.

3. Проверим правильность ранжирования

- определим расчетную сумму рангов (столбец 5):

$$\text{Расчетная сумма рангов} = 6 + 7 + 5 + 4 + 2,5 + 1 + 2,5 = 28$$

- определим правильную сумму рангов по формуле:

$$\text{Правильная сумма рангов} = \frac{n * (n + 1)}{2} = \frac{7 * (7 + 1)}{2} = \frac{56}{2} = 28$$

Ранжирование выполнено правильно.

4. Определим **нетипичные** изменения. Для этого присвоим каждому рангу знак в соответствии с направлением изменения признака: если значение признака увеличилось, то поставим перед рангом знак «+», если уменьшилось – то знак «-» (столбец 6 таблицы 5). Как видим, у пяти пациентов диурез увеличился и только у двоих уменьшился. Те изменения, которых оказалось больше, являются *типичными (частыми)*, которых меньше – *нетипичными (редкими)*. В нашем примере типичных изменений 5, нетипичных 2 (у пациентов D и H).

5. Подсчитаем сумму рангов **нетипичных** изменений. В нашем примере она равна:

$$5 + 2,5 = 7,5$$

Эта величина и будет являться эмпирическим значением критерия Уилкоксона ( $W_{\text{эмп}}=7,5$ ).

### III. Оценить эмпирическое значение критерия Уилкоксона ( $W_{\text{эмп}}$ ) и сделать вывод

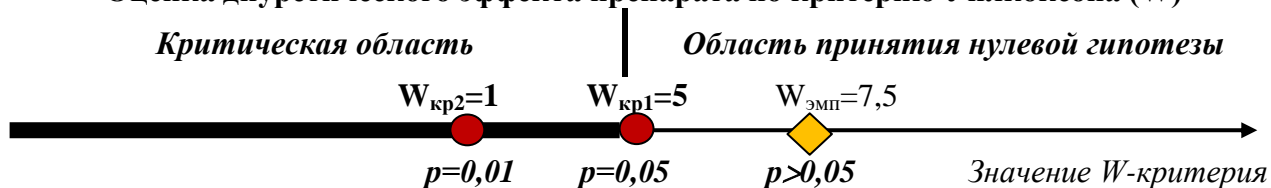
1. Определить критическое значение критерия Уилкоксона (таблица 6). Для этого необходимо:

- определить интересующий исследователя уровень значимости ( $p$ ) ( $p=0,05$  или  $p=0,01$ )
- найти критическое значение для данного объема выборки; в нашем примере для выборки из 8 единиц  $W_{\text{кр}1} = 5$  ( $p=0,05$ ) и  $W_{\text{кр}2} = 1$  ( $p=0,01$ )

2. Сравнить эмпирическое значение критерия Уилкоксона ( $W_{\text{эмп}}$ ) с его критическим значением ( $W_{\text{кр}}$ ). У критерия Уилкоксона так же, как и у критерия Манна-Уитни, критическая область располагается слева от критической точки. Чем меньше эмпирическое значение критерия Уилкоксона, тем выше статистическая значимость изменений, произошедших с единицами выборки (рисунок 2).

Рисунок 2

#### Оценка диуретического эффекта препарата по критерию Уилкоксона ( $W$ )



## Критические значения критерия Уилкоксона

n	Уровни значимости		n	Уровни значимости	
	p=0,05	p=0,01		p=0,05	p=0,01
5	0	–	28	130	101
6	2	–	29	140	110
7	3	0	30	151	120
8	5	1	31	163	130
9	8	3	32	175	140
10	10	5	33	187	151
11	13	7	34	200	162
12	17	9	35	213	173
13	21	12	36	227	185
14	25	15	37	241	198
15	30	19	38	256	211
16	35	23	39	271	224
17	41	27	40	286	238
18	47	32	41	302	252
19	53	37	42	319	266
20	60	43	43	336	281
21	67	49	44	353	296
22	75	55	45	371	312
23	83	62	46	389	328
24	91	69	47	407	345
25	100	76	48	426	362
26	110	84	49	446	379
27	119	92	50	466	397

**Вывод:** Нулевую гипотезу об отсутствии различия групп наблюдений отклонить нельзя. Эксперимент не выявил диуретический эффект лекарственного препарата ( $p > 0,05$ ).

### III. КРИТЕРИЙ ПИРСОНА $\chi^2$ (хи-квадрат)

Мы рассмотрели два непараметрических критерия, которые позволяют сравнить две совокупности по распределению количественного признака. А если изучаемый признак является не количественным, а качественным? А если сравниваемых групп не две, а больше? В этом случае можно выполнить расчет и оценку непараметрического критерия Пирсона  $\chi^2$  (читается «хи-квадрат»).

Для расчета критерия  $\chi^2$  строят так называемую таблицу сопряженности. По строкам в ней располагаются сравниваемые группы (две или более), по столбцам – интересующие значения качественного признака (два или более). Самая простая таблица сопряженности «2 x 2», она состоит из двух строк и двух столбцов (т.е. сравниваются две группы по признаку с двумя значениями). В таблице сопряженности по строкам и по столбцам обязательно подводят итоги, т.к. они участвуют в расчетах.

Данные помещают в таблицу сопряженности в форме *фактических абсолютных частот*. Каждая ячейка содержит число, которое показывает, сколько единиц наблюдения в данной группе имели данное значение признака. Каждая единица наблюдения может быть представлена в таблице сопряженности только один раз, т.е.

только в одной ячейке. Фактические частоты всегда являются натуральными числами (т.е. целыми положительными числами, полученными в результате подсчета).

Далее для каждой ячейки таблицы сопряженности, рассчитывают *ожидаемую частоту*. Она показывает, сколько единиц наблюдения в группе имели бы данное значение признака, если бы между группами различий не было. Ожидаемые частоты часто оказываются дробными числами, т.к. они получены в результате вычислений.

После расчета ожидаемых частот обязательно нужно проверить, соблюдается ли необходимое условие для последующего применения критерия  $\chi^2$ , а именно:

- в таблице «2 x 2» в каждой ячейке ожидаемая частота должна быть не менее «5».
- в таблицах большего размера не менее 80% ячеек должны содержать ожидаемую частоту «5» или более. Если же это условие не выполняется, то следует объединить соседние строки (или соседние колонки), но так, чтобы сохранить смысл данных. Затем ожидаемые частоты пересчитывают. Укрупнение ячеек следует выполнять до тех пор, пока не будет соблюдаться это условие.

Суть метода  $\chi^2$  заключается в сравнении фактических частот с ожидаемыми:

- если различие фактических и ожидаемых частот невелико, то это дает аргументы в пользу нулевой гипотезы о том, что группы наблюдений одинаковы;
- если фактические частоты существенно отличаются от ожидаемых, то, скорее всего, группы имеют статистически значимые различия.

Методику расчета и оценки критерия  $\chi^2$  рассмотрим на примере следующей задачи.

### ЗАДАЧА 3

В таблице 7 представлено распределение работников предприятия «К» по числу острых заболеваний в течение года. Требуется определить, различаются ли работники разного пола по этому признаку.

Таблица 7

#### Распределение работников предприятия «К» по кратности острых заболеваний

Пол	Число острых заболеваний за год			Итого
	Не болели	1-2 раза	3 раза и более	
1. Мужчины	29	36	15	80
2. Женщины	14	24	12	50
Оба пола:	43	60	27	130

#### Решение.

##### I. Выдвинуть нулевую и альтернативную гипотезы.

- Нулевая гипотеза ( $H_0$ ): различия мужчин и женщин, работающих на предприятии «К», по числу острых заболеваний за год статистически незначимы
- Альтернативная гипотеза ( $H_1$ ): мужчины и женщины, работающие на предприятии «К», различаются по числу острых заболеваний за год

##### II. Рассчитать эмпирическое значение критерия Пирсона $\chi^2$

1. Рассчитаем ожидаемые частоты для каждой ячейки таблицы сопряженности по формуле:

$$\frac{\text{Сумма по столбцу} * \text{Сумма по строке}}$$

Общий итог



**Мужчины:**

<b>а) Не болели</b>	<b>б) 1-2 раза</b>	<b>в) 3 раза и более</b>
$\frac{43 * 80}{130} = 26,5$	$\frac{60 * 80}{130} = 36,9$	$\frac{27 * 80}{130} = 16,6$

**Женщины:**

<b>а) Не болели</b>	<b>б) 1-2 раза</b>	<b>в) 3 раза и более</b>
$\frac{43 * 50}{130} = 16,5$	$\frac{60 * 50}{130} = 23,1$	$\frac{27 * 50}{130} = 10,4$

Для наглядности объединим фактические и ожидаемые частоты в одной таблице (таблица 8). Ожидаемые частоты выделены полужирным курсивом.

Таблица 8

**Фактические и ожидаемые абсолютные частоты заболеваемости работников предприятия «К» острыми заболеваниями**

Пол	Число острых заболеваний за год					
	Не болели		1 - 2 раза		3 раза и более	
	Фактич. частота (P)	<i>Ожидаем. частота (P')</i>	Фактич. частота (P)	<i>Ожидаем. частота (P')</i>	Фактич. частота (P)	<i>Ожидаем. частота (P')</i>
1. Мужчины	29	<i>26,5</i>	36	<i>36,9</i>	15	<i>16,6</i>
2. Женщины	14	<i>16,5</i>	24	<i>23,1</i>	12	<i>10,4</i>

Как видим фактические и ожидаемые частоты заболеваемости мужчин и женщин, работающих на предприятии «К», не слишком-то отличаются. Однако, чтобы окончательно «взвесить» аргументы против нулевой гипотезы, нам необходимо рассчитать критерий  $\chi^2$ .

2. Проверим соблюдение условия: ожидаемые частоты во всех ячейках таблицы сопряженности должны быть больше «5». Это условие выполняется (см. таблицу 8).

3. Рассчитаем критерий Пирсона  $\chi^2$  по формуле:

$$\chi^2 = \sum \frac{(P_i - P'_i)^2}{P'_i}$$

где

$P_i$  – фактическая частота

$P'_i$  – ожидаемая частота

Если таблица сопряженности самая маленькая (т.е. «2x2»), то в формулу вводят поправку:

$$\chi^2 = \sum \frac{(P_i - P'_i - 0,5)^2}{P'_i}$$

В нашем примере таблица сопряженности большего размера («2x3»), поэтому вводить поправку в формулу нам не нужно.

$$\chi^2 = \sum \frac{(P_i - P'_i)^2}{P'_i} = \frac{(29,0 - 26,5)^2}{26,5} + \frac{(36,0 - 36,9)^2}{36,9} + \frac{(15,0 - 16,6)^2}{16,6} + \frac{(14,0 - 16,5)^2}{16,5} + \frac{(24,0 - 23,1)^2}{23,1} + \frac{(12,0 - 10,4)^2}{10,4} = 0,24 + 0,02 + 0,15 + 0,38 + 0,04 + 0,25 = 1,08$$

**III. Найти критическое значение критерия  $\chi^2$  по таблице  $\chi^2$ -распределения Пирсона (таблица 9).**

- Найдем число степеней свободы (df). Для таблиц сопряженности число степеней свободы определяют по формуле:

$$df = (a - 1) * (b - 1), \text{ где}$$

a – число столбцов (без итога)

b – число строк (без итога)

Нетрудно рассчитать, что для таблиц сопряженности «2 x 2» df = 1

В нашем примере: **df = (3 - 1) \* (2 - 1) = 2**

- Выбираем уровень значимости **p=0,05**
- По таблице  $\chi^2$ -распределения Пирсона находим критическое значение  $\chi^2$  для df=2.  **$\chi^2_{кр}=5,991$**

Таблица 9

**Критические значения критерия  $\chi^2$  Пирсона**

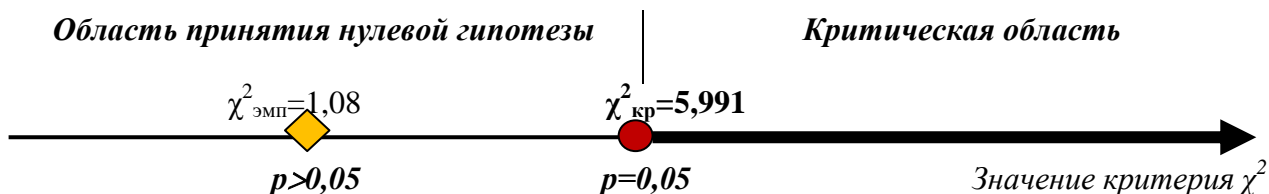
df	Двустороннее значение «p»			
	0,10	0,05	0,01	0,001
1	2,706	3,841	6,635	10,827
<b>2</b>	4,605	<b>5,991</b>	9,210	13,815
3	6,251	7,815	11,345	16,266
4	7,779	9,488	13,277	18,466
5	9,236	11,070	15,086	20,515
6	10,645	12,592	16,812	22,457
7	12,017	14,067	18,475	24,321
8	13,362	15,507	20,090	26,124
9	14,684	16,919	21,666	27,877
10	15,987	18,307	23,209	29,588
11	17,275	19,675	24,725	31,264
12	18,549	21,026	26,217	32,909
13	19,812	22,362	27,688	34,527
14	21,064	23,685	29,141	36,124
15	22,307	24,996	30,578	37,698
20	28,412	31,410	37,566	45,314
25	34,382	37,652	44,314	52,619
30	40,256	43,773	50,892	59,702
40	51,805	55,758	63,691	73,403
50	63,167	67,505	76,154	86,660
60	74,397	79,082	88,379	99,608
70	85,527	90,531	100,43	112,32
80	96,578	101,88	112,33	124,84
90	107,57	113,15	124,12	137,21
100	118,50	124,34	135,81	149,45

#### IV. Сравнить эмпирическое значение критерия $\chi^2$ с его критическим значением и сделать выводы

У критерия Пирсона  $\chi^2$  критическая область располагается справа от критической точки, как и у большинства статистических критериев. Чем больше эмпирическое значение  $\chi^2$ , тем больше вероятность того, что сравниваемые группы действительно различаются (рисунок 3).

Рисунок 3

#### Оценка различия мужчин и женщин, работающих на предприятии «К», по числу острых заболеваний за год, используя критерий $\chi^2$ Пирсона



**Вывод:** Аргументов для отклонения нулевой гипотезы недостаточно. Между мужчинами и женщинами, работающими на предприятии «К», статистически значимых различий в числе острых заболеваний за год не выявлено ( $p>0,05$ ).

#### IV. КРИТЕРИЙ МАКНЕМАРА

Критерий МакНемара используется для сравнения двух связанных (зависимых) групп по альтернативному описательному признаку. Напомним, что этот признак имеет только два противоположных значения – «да» и «нет». Каждая единица наблюдения изучается исследователем в двух состояниях. Интересующий описательный признак в том и в другом состоянии мог либо присутствовать, либо отсутствовать.

Таким образом, каждая единица наблюдения оказывается отнесенной к одному из четырех классов:

1. Изучаемое свойство имелось в обоих состояниях
2. Изучаемое свойство отсутствовало в обоих состояниях
3. Изучаемое свойство присутствовало в первом состоянии, но отсутствовало во втором
4. Изучаемое свойство отсутствовало в первом состоянии, но присутствовало во втором

Данные представляют в форме таблицы сопряженности «2x2» (таблица 10). Рабочие ячейки этой таблицы заполняются **фактическими абсолютными частотами**. Каждая единица наблюдения может быть представлена только в одной ячейке. Для наглядности по строкам и по столбцам можно подвести итоги, но в расчете критерия МакНемара эти итоги не участвуют.

Таблица сопряженности для расчета критерия МакНемара

		Первое состояние		Общее число наблюдений:
		Свойство имеется	Свойство отсутствует	
Второе состояние	Свойство имеется	W	X	W+X
	Свойство отсутствует	Y	Z	Y+Z
Общее число наблюдений:		W+Y	X+Z	n=W+X+Y+Z

В статистическом анализе не учитываются те единицы наблюдения, у которых никаких изменений не произошло:

- Интересующее свойство присутствовало в обоих состояниях (величина из ячейки W)
- Интересующее свойство отсутствовало в обоих состояниях (величина из ячейки Z)

Для расчета критерия МакНемара используют те единицы наблюдения, у которых интересующее свойство присутствовало только в одном состоянии (величины из ячеек X и Y).

Критерий МакНемара подчиняется распределению  $\chi^2$  с одной степенью свободы (df=1) и рассчитывается по формуле:

$$\chi^2 = \frac{(|X - Y| - 1)^2}{X + Y}$$

Рассмотрим применение критерия МакНемара на следующем примере.

#### Задача 4

Сравнили два метода выявления кариеса – визуальное обследование и радиографический метод. В исследовании участвовало 100 пациентов, у которых проверяли состояние первого большого коренного зуба. Полученные данные представлены в таблице 11.

Таблица 11

#### Результаты выявления кариеса при визуальном обследовании и с применением радиографического метода

		Радиографический метод		Общее число наблюдений:
		Кариес есть	Кариеса нет	
Визуальное обследование	Кариес есть	45	4	49
	Кариеса нет	17	34	51
Общее число наблюдений:		62	38	100

Как видим, визуальное обследование выявило кариес почти у половины испытуемых (49%), а радиографический метод – почти у двух третей (62%). Но, быть

может, обнаруженные различия двух методов диагностики кариеса – случайны, связаны только с данной конкретной выборкой? Можно ли считать эти два метода диагностики кариеса равнозначными или же нет?

### Решение

#### I. Выдвигаем нулевую и альтернативную гипотезы

- *Нулевая гипотеза ( $H_0$ ):* визуальное обследование и радиографический метод диагностики кариеса выявляют примерно одинаковое число больных лиц
- *Альтернативная гипотеза ( $H_1$ ):* визуальное обследование и радиографический метод принципиально различаются по возможностям выявления кариеса

#### II. Рассчитаем значение критерия МакНемара по эмпирическим данным

$$\chi^2_{\text{эмп}} = \frac{(|X - Y| - 1)^2}{X + Y} = \frac{(17 - 4 - 1)^2}{17 + 4} = \frac{144}{21} = 6,86$$

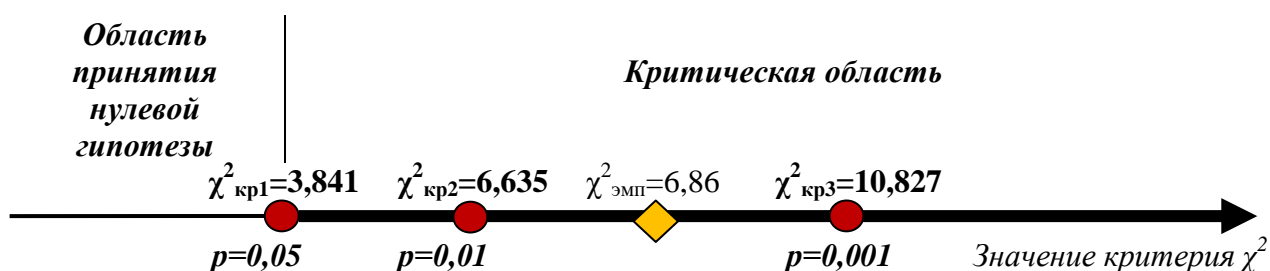
#### III. Сравним эмпирическое значение критерия МакНемара с его критическим значением и сделаем выводы

По таблице  $\chi^2$ -распределения Пирсона (таблица 9) находим критическое значение  $\chi^2$  для df=1:

- если  $p=0,05$ , то  $\chi^2_{\text{кр1}}=3,841$
- если  $p=0,01$ , то  $\chi^2_{\text{кр2}}=6,635$
- если  $p=0,001$ , то  $\chi^2_{\text{кр3}}=10,827$

Рисунок 4

#### Оценка различия двух методов диагностики кариеса по критерию МакНемара



**Вывод:** Нулевую гипотезу о равенстве возможностей двух методов диагностики кариеса (визуальное обследование и радиографический метод) можно отклонить с уровнем значимости  $p < 0,01$ . С уверенностью более 99% можно утверждать, что различие сравниваемых диагностических методов статистически значимо: радиографический метод чаще обнаруживает кариес, чем визуальное обследование.

### ОШИБКИ, ВОЗНИКАЮЩИЕ В ПРОЦЕССЕ ПРОВЕРКИ ГИПОТЕЗ

В прошлом семестре мы познакомились с двумя видами ошибок, которые могут присутствовать в статистическом исследовании. Кратко напомним, что речь идет о случайной ошибке и систематической ошибке.

**Случайная ошибка** обусловлена разнообразием изучаемого признака. В генеральной совокупности присутствуют единицы наблюдения с самыми разными значениями интересующего исследователя признака. Какие единицы попадут в выборку, с какими значениями признака – это полная случайность. Исследователь стремится сделать свою выборку *репрезентативной*, т.е. представительной, способной правильно охарактеризовать изучаемую генеральную совокупность. Для этого исследователь

формирует выборку достаточно большого объема (т.е. заботится о *количественной репрезентативности*), а также старается сделать так, чтобы структура выборки по самым важным для данного исследования факторным признакам была бы идентична структуре генеральной совокупности (т.е. заботится о *качественной репрезентативности*). Из материала этого семестра вы знаете, что величину случайной ошибки (т.е. стандартную ошибку) можно рассчитать по специальным формулам. Величина этой ошибки формирует доверительный интервал. При выборочном исследовании случайная ошибка присутствует всегда.

**Систематическая ошибка** появляется в результате каких-либо дефектов в организации и проведении исследования. По возможности исследователь старается не допустить ее появления: он самым тщательным образом продумывает дизайн своей работы, аккуратно (не «наобум») отбирает единицы наблюдения, внимательно собирает, обрабатывает и анализирует материал. Но, к сожалению, далеко не всегда это помогает предотвратить систематические ошибки. Например, при сборе материала методом опроса следует помнить о возможном стремлении респондентов скрыть какую-либо информацию или представить ее в искаженной форме.

Эти ошибки могут привести к тому, что, работая с научной гипотезой, исследователь может сделать неверный вывод. Дело в том, что

- принятие гипотезы еще не означает, что она действительно истинна,
- отклонение гипотезы еще не означает, что она действительно ложна.

Научное исследование – это шаг из света в темноту, в непознанное, в неведомое. Представьте, что вы заходите в темную комнату. Вы вполне можете не разглядеть многое из того, что там имеется, либо наоборот – вам может примерещиться что-то, чего там на самом деле нет.

Для пояснения этой проблемы приведем слова древнегреческого философа-стоика I-II века Эпиктета (ок.50-ок.138):

*«Для разума при восприятии окружающего возможны четыре ситуации:*

- *явление существует и воспринимается как оно есть,*
- *явление не существует и не воспринимается,*
- *явление существует, но не воспринимается,*
- *явление не существует, но кажется существующим*

*Правильно распознать каждый случай – задача, достойная мудреца.»*

В качестве образного примера приведем ситуацию суда.

Таблица 12

### Решение, принимаемое судом

		ИСТИНА	
		Виновен	Невиновен
ВЫВОДЫ СУДА	Виновен	Справедливо осужден	Ошибочно осужден
	Невиновен	Ошибочно оправдан	Справедливо оправдан

Истина – виновен подсудимый или же нет – судье не известна. По итогам заседания судья должен либо осудить человека, либо оправдать. И в том, и в другом случае судья может оказаться прав, но может и ошибиться:

1. Подсудимый виновен, и судья признал его виновным («явление существует и воспринимается» по Эпиктету)

2. Подсудимый невиновен, и судья оправдал его («явление не существует и не воспринимается»)
3. Подсудимый на самом деле виновен, но судьей ошибочно оправдан («явление существует, но не воспринимается»)
4. Подсудимый на самом деле невиновен, но судья ошибочно признал его виновным («явление не существует, но кажется существующим»)

С помощью такой же матрицы можно представить выводы исследователя. Предположим, что речь идет об экспериментальном исследовании, где изучается эффективность нового медицинского вмешательства по сравнению с современной типичной практикой.

Таблица 13

**Выводы исследователя по итогам проверки научной гипотезы**

		ИСТИНА	
		Новое вмешательство лучше	Новое вмешательство такое же
ВЫВОДЫ ИССЛЕДОВАТЕЛЯ	Новое вмешательство лучше (группы различаются)	Новое вмешательство действительно лучше	Ошибка первого рода (ошибка $\alpha$ )
	Новое вмешательство такое же (группы одинаковы)	Ошибка второго рода (ошибка $\beta$ )	Новое вмешательство действительно такое же

Какова истина – более эффективно новое вмешательство или же оно не имеет преимуществ перед типичной практикой – исследователю не известно. Для выяснения истины он выполняет научное исследование. Он выдвигает нулевую гипотезу о том, что сравниваемые группы одинаковы, т.е. новое вмешательство примерно настолько же эффективно, как современная практика. По результатам экспериментальной проверки исследователь может принять свою нулевую гипотезу, либо же отвергнуть ее. И в том, и в другом случае он может быть прав, но может и ошибиться.

1. Новое вмешательство действительно лучше, и исследователь экспериментальным путем это подтвердил. Он выявил статистически значимые различия сравниваемых групп, правильно отклонил нулевую гипотезу, которая действительно была ложной («явление существует и воспринимается»)
2. Новое вмешательство по уровню эффективности примерно такое же, как типичная современная практика, и исследователь в эксперименте это увидел. Нулевая гипотеза о том, что группы одинаковы, является истинной, и исследователь совершенно правильно с нулевой гипотезой согласился («явление не существует и не воспринимается»).
3. Новое вмешательство на самом деле лучше, эффективнее типичной практики, но исследователю доказать это не удалось. Нулевая гипотеза о том, что сравниваемые

группы одинаковы, на самом деле является ложной, однако исследователь не смог найти достаточно сильных статистических аргументов, чтобы ее отклонить, и вынужден был с ней согласиться. Быть может, выборка была слишком мала (отсутствовала количественная репрезентативность), или физиологические характеристики испытуемых были какими-то «нетипичными» (отсутствовала качественная репрезентативность), но, так или иначе исследователь «оказался в черных очках слепого» – он не увидел то, что есть («явление существует, но не воспринимается» по Эпиктету). Такая ошибка называется **«ошибка второго рода», «ошибка типа β»**.

4. Новое вмешательство по своей эффективности примерно такое же, как типичная современная практика. Нулевая гипотеза о том, что сравниваемые группы одинаковы, является истинной. Однако в силу каких-то причин (опять же связанных с отсутствием количественной и/или качественной репрезентативности выборки) исследователь получил данные, позволившие ему отклонить эту истинную нулевую гипотезу. Он увидел «миражи», увидел несуществующее различие групп и сделал вывод о более высокой эффективности нового медицинского вмешательства («явление не существует, но кажется существующим»). Такая ошибка называется **«ошибка первого рода», «ошибка типа α»**.

Последствия ошибочных выводов исследователя могут быть разные.

- Если ошибочно будет отвергнут новый более эффективный метод лечения (т.е. совершена ошибка второго рода, ошибка β), то значит будут использоваться другие, менее эффективные методы.
- Если будет ошибочно доказана высокая эффективность какого-либо вмешательства, а высокой эффективностью на самом деле это вмешательство не обладает (т.е. допущена ошибка первого рода, ошибка α), то применение неправильного метода лечения может нанести ущерб здоровью пациента и ресурсам здравоохранения.

С **ошибкой первого рода (α)** вы знакомы уже давно, только знали ее под другим названием – уровень значимости (p). Еще раз напомним, что вероятность допустить эту ошибку в медико-биологических исследованиях должна быть **не более 5% (или 0,05)**. Другими словами, вероятность увидеть различия там, где их нет, не должна превышать 5%. Уверенность в том, что сравниваемые группы действительно различаются должна быть 95% и выше. Чем меньше уровень значимости – тем лучше, тем надежнее полученные выводы.

Когда ведут речь об **ошибке второго рода (β)**, то часто говорят не о значении самой этой ошибки, а о величине, которая с ней связана – **мощность (чувствительность) исследования**. Мощность исследования рассчитывается по формуле:

$$\text{Мощность исследования} = 1 - \beta$$

Другими словами, мощность исследования – это вероятность **НЕ ДОПУСТИТЬ** ошибку второго рода (β), **вероятность выявить эффект вмешательства, если он на самом деле существует**, вероятность увидеть статистически значимые различия групп, если они действительно есть. Достичь идеальной 100% мощности при выборочном исследовании невозможно, как бы хорошо это исследование ни было спланировано и выполнено. Всегда будет сохраняться хоть какая-то вероятность, что в силу некоторых причин существующий эффект вмешательства не будет выявлен.

Мощность задается до начала исследования и должна быть **не менее 70% (0,7)**. То есть, если вмешательство действительно эффективно, то у исследователя должна быть как минимум 70%-ая уверенность увидеть этот эффект.



Можно также говорить о мощности (чувствительности) статистических критериев. Параметрические критерии обладают большей мощностью, чем непараметрические. Если сравниваемые группы действительно различаются, то с помощью t-критерия Стьюдента легче отклонить нулевую гипотезу, чем с помощью его непараметрических аналогов. Поэтому специалисты призывают исследователей по возможности применять параметрические методы, и только тогда, когда не соблюдаются обязательные условия их применения, использовать методы непараметрической статистики.

Ошибки первого и второго рода ( $\alpha$  и  $\beta$ ) взаимосвязаны: уменьшение одной из них приводит к увеличению другой. Одновременно уменьшить обе ошибки можно только путем увеличения объема исследования.