

Пример 2.4. Дана таблица экспериментальных данных. Проверить нулевую гипотезу о равенстве признака в этих выборках на уровне значимости $\alpha = 0,05$.

При несоответствии данных какому-либо закону распределения в практике исследований используются *непараметрические критерии статистики*, в частности, критерий Манна-Уитни, применение которого хотя и не требует большого объема выборки и знания вида распределения, но, тем не менее, зависит от ряда условий (см. п.2.2.1). Критерий предназначен для оценки различий между двумя выборками (скажем X и Y) по уровню какого-либо количественно измеренного признака, и что более важно, критерий Манна-Уитни позволяет оценивать выборки при неизвестном законе распределения признака (общепринятая интерпретация – проверка равенства медиан). Тест U позволяет проверить, существует ли достоверная разница между двумя независимыми выборками *после того*, как сгруппированные данные этих выборок ранжированы и вычислены суммы рангов для каждой выборки. Чем меньше перекрещивающихся значений (чем меньше U), тем более вероятно, что различия достоверны: т.е.

если $U < U_{кр}$, то нулевая гипотеза отвергается.

X	Y
14	18
20	19
15	22
11	17
16	24
13	21
16	25
19	26
15	24
9	15
	18
	22

При построении решения алгоритм предусматривает следующее.

1. В диапазон A5:A14 заносятся исходные данные по выборке X, в B5:B16 – по выборке Y. В соответствующие ячейки (рис. А) заносятся поясняющие данные.
2. В ячейку F4 заносится величина уровня значимости.
3. В ячейках F5, F6 формулами =СЧЁТ(A5:A14) {=COUNT(A5:A14)} и =СЧЁТ(B5:B16) {=COUNT(B5:B16)} подсчитываются размеры выборок.
4. Для удобства подсчета рангов определим массив XY данных объединенных выборок – выделим диапазон A5:B16, далее ВСТАВКА–ИМЯ–ПРИСВОИТЬ {ВСТАВКА–НАЗВАНИЯ–ОПРЕДЕЛИТЬ} имя XY.
5. В ячейку C5 вводится формула =(СЧЁТ(XY) + 1 - РАНГ(A5; XY; 0) + РАНГ(A5; XY; 1)) /2 {(COUNT(XY)+1 -RANK(A5;XY;0) +RANK(A5; XY;1))/2} и автозаполнением тиражируется по диапазонам A6:A14 и B5:B16.
6. В ячейках F8, F9 формулами =СУММ(C5:C14) {=SUM(C5:C14)} и =СУММ (D5:D16) {=SUM(D5:D16)} подсчитываются суммы рангов по выборкам.
7. В ячейке F11 формулой =ЕСЛИ(F8>F9;F5;F6) {=IF(F8>F9;F5;F6)} определяется размер выборки с максимальной суммой рангов, а в ячейку F12 формулой =ЕСЛИ(F8>F9; F8;F9) {=IF(F8>F9; F8;F9)} заносится эта максимальная сумма.

	A	B	C	D	E	F	G	H	I	J
1	Нулевая гипотеза: среднеарифметические двух									
2	выборок не различаются									
3					$=(СЧЁТ(ХУ)+1-РАНГ(В5;ХУ;0)+РАНГ(В5;ХУ;1))/2$					
4	X	Y	R _x	R _y	$\alpha =$	0,05				
5	14	18	4,0	11,5	$n_x =$	10	←	=СЧЁТ(A5:A14)		
6	20	19	15,0	13,5	$n_y =$	12	←	=СЧЁТ(B5:B16)		
7	15	22	6,0	17,5						
8	11	17	2,0	10,0	$\Sigma R_x =$	67,5	←	=СУММ(C5:C14)		
9	16	24	8,5	19,5	$\Sigma R_y =$	185,5	←	=СУММ(D5:D16)		
10	13	21	3,0	16,0						
11	16	25	8,5	21,0	$n^* =$	12	←	=ЕСЛИ(F8>F9;F5;F6)		
12	19	26	13,5	22,0	$T^* =$	185,5	←	=ЕСЛИ(F8>F9;F8;F9)		
13	15	24	6,0	19,5						
14	9	15	1,0	6,0	$U =$	12,5	←	=F5*F6+F11*(F11+1)/2-F12		
15		18		11,5	$U_{кр} =$	34				
16		22		17,5						
17										
18	Поскольку $U < U_{кр}$ то нулевая гипотеза отвергается									
19	на выбранном уровне значимости									
20	Различия между выборками признаются									
21	статистически значимыми									

Рис. А. Скриншот расчетного листа MS Excel для примера 2.4.

8. В ячейку F14 вводится формула =F5*F6+F11*(F11+1)/2-F12, реализующая расчет статистики Манна-Уитни по формуле

$$U = (n_x \cdot n_y) + \frac{n_*(n_* + 1)}{2} - T_*$$

9. Выбором из таблицы (см. **таблица $U_{кр}$**) в ячейку F15 вводится критическое значение статистики Манна-Уитни.

10. Из сравнения величин критериев ($U < U_{кр}$) следует возможность принятия альтернативной гипотезы о достоверных различиях средних арифметических.

Правила ранжирования.

1. Меньшему значению начисляется меньший ранг, как правило, это. Наибольшему значению начисляется ранг, соответствующий количеству ранжируемых значений (если $n=10$, то наибольшее значение получит ранг 10). Исключения оговорены в пункте 2.
2. Если несколько значений равны, им начисляется ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны. Например, если в упорядоченной по возрастанию выборке три следующие за минимальным значением величины равны, то каждой из них присваивается ранг

$$\frac{2 + 3 + 4}{3} = \frac{9}{3} = 3.$$

3. Общая сумма рангов должна совпадать с расчетной, которая определяется по формуле

$$\sum (R_i) = \frac{n(n + 1)}{2}.$$

где n – общее количество ранжируемых значений. Несовпадение реальной и расчетной сумм рангов будет свидетельствовать об ошибке, допущенной при начислении рангов или их суммировании.