Лекция 8

ОСНОВНЫЕ НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ, ПРИМЕНЯЕМЫЕ ПРИ АНАЛИЗЕ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЯ ДИАГНОСТИЧЕСКОЙ ИНФОРМАЦИИ

К непараметрическим методам проверки гипотез относят методы, не использующие допущения о виде распределения данных измерений и определении его параметров. Это является преимуществом этих методов. Их недостаток — меньшая по сравнению с параметрическими методами мощность. Применение большинства непараметрических критериев требует небольшого объема вычислений и их удобно применять для быстрого опровержения нулевой гипотезы, например отличия полученных результатов от ранее известных.

1. Критерий знаков для определения равенства средних значений двух выборок

Одним из наиболее простых непараметрических критериев для принятия решения о различии выборок является критерий знаков. В основном он применяется для сравнения двух методов, процессов, приборов и т. д., когда результаты измерений попарно связаны, образуя зависимые, или связанные выборки.

Основой *критерия знаков* является факт, что при одинаковых объемах неразличающихся выборок с равной вероятностью следует ожидать появления положительных и отрицательных разностей между результатами отдельных пар измерений. Заметное отклонение от середины есть свидетельство различия средних значений в двух выборках, а точнее — медиан их распределений.

Для практического применения этого критерия результаты, полученные для двух совокупностей данных, записывают попарно в виде таблицы, затем определяют знаки разностей пар с одинаковыми индексами, вычитая из верхнего значения нижнее, и проставляют знаки в последней строке. Нулевые разности вычеркивают.

Первая выборка	x_1	x_2	…	x_N
Вторая выборка	y_1	y_2	…	y_N
Знак разности $x_i - y_i$	$x_1 - y_1$	$x_2 - y_2$	…	$x_N - y_N$

Пусть проведено N пар испытаний, в которых получено k положительных разностей, m отрицательных, l нулевых $\left(N = k + m + l\right)$. Обозначим через $n_1 = N - l$ число результатов, оставшихся после отбрасывания пар с нулевыми разностями. Нулевая гипотеза о равенстве медиан не отвергается, если k и m попадают в допустимый интервал значений для выбранного уровня значимости и полученного количества ненулевых разностей n_1. Допустимый интервал определяют по статистическим таблицам. Если k или m (или оба числа) выходят за допустимые пределы, нулевую гипотезу отвергают, то есть медианы выборок значимо различаются, следовательно, различаются и сами выборки.

Если альтернативной гипотезой является предположение о том, что медианы просто различны, используют двусторонний критерий. Если же альтернативная гипотеза состоит в том, что медиана первой выборки больше, чем медиана второй (или наоборот), необходимо применить односторонний критерий.

Для иллюстрации применения критерия знаков вновь обратимся к примеру Л5.2 (лекция 5). Пример удобен тем, что его анализ позволяет обойтись простейшими вычислительными средствами и вместе с тем затрагивает ряд характерных для такого анализа проблем. Прежде всего, сформулируем задачу в терминах статистики.

Имеются две выборки данных: первая — для труб с покрытием, вторая — для труб без покрытия. На основе этих выборок требуется принять одну из двух взаимоисключающих гипотез. Первая (основная) H_0 состоит в том, что выборки статистически не различаются, то есть, что те и другие трубы имеют одинаковую коррозионную стойкость.

Вторая (альтернативная) гипотеза H_1 заключается в том, что выборки различаются, то есть трубы во второй выборке отличаются от первой по коррозионной стойкости. Обратим внимание на то, что альтернативная гипотеза предусматривает коррозионную стойкость труб из второй выборки не как более низкую, а как отличающуюся от стойкости первой партии. Таким образом, отражена корректная постановка задачи, согласно которой учтена маловероятная, но не невозможная ситуация, когда применение покрытия ухудшает стойкость труб. К такому результату может привести неудачный выбор покрытия или технологии его нанесения. Таким образом, сформулирована задача: на основании сравнения двух выборок определить, какую из двух гипотез — H_0 или H_1 следует принять.

Применим для решения задачи критерий знаков. В соответствии с рассмотренной выше методикой построим таблицу разностей:

Трубы с покрытием	39	43	43	52	52	59	40	45	47	62	40	27
Трубы без покрытия	42	37	61	74	55	57	44	55	37	70	52	55
Знак разности	-	+	-	-	-	+	-	-	+	-	-	-

Нулевые разности отсутствуют, следовательно, $l = 0$, $n_1 = n = 12$, $k = 3$, $m = 9$.

Если принять уровень значимости $\alpha = 0{,}1$, допустимый интервал значений изменения k и m для принятия нулевой гипотезы согласно статистическим таблицам составляет от 3 до 9. Следовательно, нет достаточных оснований для того, чтобы считать выборки различающимися.

К сожалению, применение непараметрических критериев связано с трудностями оценки их мощности, и поэтому практически ничего нельзя сказать о вероятности того, что данное решение является ошибочным. Это естественная плата за простоту критерия и ограниченный объем выборки. Если бы выборки различались, например, было бы получено $k = 2$ и $m = 10$, можно было бы сказать, что с вероятностью ошибки, не превышающей 10%, выборки различаются. Заметим, что отклонение гипотезы о различии медиан еще не означает неразличимости выборок.

2. Ранговый критерий для сравнения дисперсий (критерий Сиджела-Тьюки)

Отсутствие оценки мощности критерия при неразличимости выборок побуждает к применению других методов различения (распознавания), так как равенство средних значений еще не гарантирует полного сходства выборок. Показателен следующий пример. На одном из заводов, выпускающих электрооборудование для АЭС, возник вопрос, почему валы двигателей одного из заводов-поставщиков служат в среднем дольше, чем валы другого. Размеры находятся в пределах допусков, поставщик металла один и тот же.

Оказалось, что при одинаковых средних размерах их разброс на первом заводе был значительно меньше допустимых, а на втором — еле вписывался в технологические нормы. Поэтому следующий этап в попытке различить выборки — сравнение степени рассеяния значений в них, то есть определение дисперсии и связанного с ней среднеквадратичного отклонения результатов от среднего. Одним из простейших непараметрических критериев сравнения дисперсий является ранговый *критерий Сиджела-Тьюки*.

Ранговые критерии занимают важное место среди непараметрических методов. Они основаны на ранговой последовательности измеренных значений величин и на расчетах с помощью ранговых чисел, значительно упрощающих расчеты. Естественно, что при переходе от измеренных значений к ранговым числам часть информации теряется, вследствие чего ранговые критерии, как и любые непараметрические критерии вообще, менее эффективны, чем параметрические.

Критерий Сиджела-Тьюки применим при равенстве характеристик положения (средних значений, медиан), размеры выборок N_1 и N_2 могут различаться, но размер каждой выборки должен быть не меньше 10, для определенности выборки нумеруют таким образом, чтобы вторая выборка была по объему не меньше первой, т. е. $N_1 \leq N_2$.

Обе выборки объединяют в единый вариационный ряд, отметив принадлежность каждого члена ряда той или иной выборке. Затем члены полученного ряда ранжируют следующим образом. Присваивают ранг R, равный 1, первому, минимальному члену ряда, ранг 2 — последнему, максимальному. Ранг 3 присваивают максимальному из оставшихся членов ряда, ранг 4 — минимальному из оставшихся и т. д. В случае, если несколько членов ряда равны друг другу, им присваивают одинаковый ранг, равный их среднему арифметическому.

Затем вычисляют сумму рангов каждой выборки R_1 и R_2. Для определенности через R_1 обозначают сумму рангов для меньшей по объему выборки, а в случае равенства объемов $\left(N_1 = N_2\right)$ за R_1 принимают меньшую сумму, то есть $R_1 < R_2$. Для устранения возможных ошибок расчетов проверяют равенство

$$R_1 + R_2 = \left(1/2\right)\left(N_1 + N_2\right)\left(N_1 + N_2 + 1\right). \qquad \text{(Л8.1)}$$

Наконец, рассчитывают статистику

$$z = \frac{\left|R_1 - 1/2 N_1\left(N_1 + N_2 + 1\right)\right| - 1/2}{\sqrt{N_1 N_2\left(N_1 + N_2 + 1\right)/12}} \qquad \text{(Л8.2)}$$

и сравнивают ее значение с критическим значением квантили нормированного нормального распределения $z_{1-\alpha/2}$. Если $z > z_{1-\alpha/2}$ принимается гипотеза о том, что дисперсии статистически различаются.

Для примера Л5.2 (лекция 5) получим таблицу рангов 1. Из нее в частности следует, что $R_1 + R_2 = 300$ (для точного подсчета нужно взять простые дроби вместо десятичных). Стоящая в правой части равенства (Л8.1) величина равна 24 · 25 / 2 = 300, то есть проверка показывает отсутствие ошибок в расчетах рангов. Статистика z равна 0,93. Из статистических таблиц находим $z_{0,2} = 0,84$, $z_{0,1} = 1,28$. Следовательно, уровень значимости α, соответствующий $z_\alpha = 0,93$, находится между 0,1 и 0,2, более точно — около 0,18. Таким образом, с вероятностью ошибки 18% можно считать выборки различающимися по дисперсии.

Таблица Л8.1. **Таблица рангов**

x_i	№ выборки	Исходные ранги	Ранги первой выборки	Ранги второй выборки
27	1	1	1	1
37	2	4	4,5	4,5
37	2	5	4,5	4,5
39	1	8	8	8
40	1	9	10,5	10,5
40	1	12	10,5	10,5
42	2	13	13	13
43	1	16	16,5	16,5
43	1	17	16,5	16,5
44	2	20	20	20
45	1	21	21	21
47	1	24	24	24
52	1	23	21,33	21,33
52	1	22	21,33	21,33
52	2	19	21,33	21,33
55	2	18	15,67	15,67
55	2	15	15,67	15,67
55	2	14	15,67	15,67
57	2	11	11	11
59	1	10	10	10
61	2	7	7	7
62	1	6	6	6
70	2	3	3	3
74	2	2	2	2
			R_1=166,66	R_2=133,33

В данном примере мы имеем дело с так называемыми зависимыми, или связанными выборками, так как индивидуальные результаты измерений попарно взаимосвязаны. В этом случае можно рассчитывать на получение статистически более надежных результатов, если учесть эту парную обусловленность, анализируя не хаотически представленные данные, а разности между значениями регистрируемых параметров. Если обозначить через d_i разности скоростей коррозии защищенного и незащищенного образцов i-й пары, то вместо двух выборок получим одну:

№ пары	1	2	3	4	5	6	7	8	9	10	11	12
Разность скоростей коррозии d, мм/год	3	–6	18	22	3	–2	4	10	–10	8	12	28

Расчет дает $\overline{d} = 7,5$; $s = 11,3$; $h_k = 1,50$; $S_{k\ выб} = 0,234$; $E_{k\ выб} = -1,096$; $t = 1,65$.

Для проверки значимости различия средних значений двух выборок необходимо убедиться, что среднее значение разностной выборки значимо отлично от нуля, то есть, что 7,5 не является нулем с точки зрения статистики. Предположим, что распределение величин в выборке подчиняется нормальному закону, что, вообще говоря, должно быть проверено, см. лекцию 7 п.1, и воспользуемся критерием равенства средних двух выборок, см. лекцию 7 п. 4.

Зададимся доверительной вероятностью $p = 1 - \alpha$ интервальной оценки *d*, равной 0,9. Согласно п. 4 лекции 9 необходимо найти значение $t_{\alpha,\, N-1}$ статистики Стьюдента. По статистическим таблицам находим $t_{0{,}1,\, 11} \approx 1{,}80$, тогда

$$z = \frac{t_{\alpha,\, N-1}}{\sqrt{2(N-1)}} s = 1{,}8 \cdot 11{,}3/4{,}69 \approx 4{,}34 .$$

Следовательно, с вероятностью 90% среднее значение лежит в пределах $\bar{x} = 7{,}5 \pm 4{,}3$ или $3{,}2 < \bar{x} < 11{,}8$.

Как видим, нуль в указанный интервал при данном уровне значимости не попадает, то есть средние значения двух выборок различаются значимо. Обратим внимание на результат, отличный от полученного с помощью критерия знаков. Критерий знаков не позволил выявить различия между средними значениями, однако заключение было осторожным: нет оснований считать различие средних значений значимым. Таким образом, как и следовало ожидать, параметрический критерий Стьюдента оказался более мощным, чем непараметрический критерий знаков, и позволил выявить различие выборок.

3. Критерий проверки соответствия эмпирической функции распределения некоторой теоретической функции (критерий Колмогорова-Смирнова)

Повышение мощности критерия может быть достигнуто за счет анализа нескольких характеристик распределения. Один из наиболее широко известных и применяемых на практике *критерий Колмогорова-Смирнова* использует распределение максимального отклонения «накопленной частости» — эмпирической функции распределения от теоретической функции распределения $F(x)$. Критерий используют для проверки соответствия эмпирической функции распределения некоторому теоретическому, как правило, нормальному. Для этого по экспериментальным данным находят среднее значение

$\bar{x}$ и дисперсию s^2 выборки и проверяют, соответствует ли выборка нормальному распределению $N\left(\bar{x},\ s^2\right)$. Для применения критерия вычисляют статистику

$$D_N = \max\ \left\{D_N^+,\ D_N^-\right\},$$

где $D_N^+ = \max\ \left\{i/N - F(x_i)\right\}$, $D_N^- = \max\ \left\{F(x_i) - (i-1)/N\right\}$. А.Н. Колмогоров доказал, что для любой непрерывной функции распределения $F(x)$ при $N \to \infty$

$$\lim_{N\to\infty} P\left\{\sqrt{N} D_N \le x\right\} = K(x) = \sum_{k=-\infty}^{\infty} (-1)^k\, \mathrm{e}^{-2k^2x^2},$$

откуда следует *непараметрический критерий Колмогорова*:

$$1 - K(k_\alpha) \approx \alpha; \qquad (N \ge 20...25),$$

где k_α — α-квантиль распределения $K(x)$.

Если $\sqrt{N} D_N > k_\alpha$, гипотеза о соответствии выборки распределению с функцией распределения $F(x)$ отклоняется с вероятностью ошибки первого рода, равной α. В приведенном виде критерий может применяться при размере выборки $N > 20$.

Для проверки гипотезы о том, что две выборки получены из одного и того же распределения, может быть применен *критерий Смирнова*, являющийся модификацией критерия Колмогорова. Для двух выборок с размерами N_1 и N_2 строят статистику

$$D_{12} = \max\left(F_{1i}^*(x) - F_{2i}^*(x)\right),$$

при превышении которой критического значения, указанного в статистических таблицах, гипотеза о принадлежности обеих выборок одному и тому же распределению отвергается.