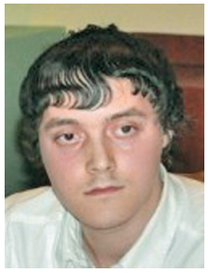


***Нинчук В.С., Чепурко В.А.***

# О ПРОВЕРКЕ СЛУЧАЙНОСТИ ДАННЫХ СО СВЯЗЯМИ

*Статья посвящена исследованию различных непараметрических методов проверки гипотезы случайности (отсутствие тренда) выборочных данных в ситуации, когда эти данные содержат повторяющиеся наблюдения. Повторяющиеся наблюдения, называемые связями, приводят к неоднозначному ранжированию. Обычно элементам связанной группы присваивается среднеарифметический ранг. В результате изменяются статистические решающие правила, поскольку статистика критерия меняет свое распределение. Настоящая статья посвящена исследованию мощности различных критериев при наличии связей.*

***Ключевые слова: с****вязь, связанная группа, ранг, инверсия, критерий Спирмена, критерий Кендалла, интегральный критерий.*

## Введение

Для получения корректных оценок и статистического прогнозирования ресурса различных сложных технических систем необходимо использовать хорошо обусловленные методы. При анализе статистических данных об отказах восстанавливаемых объектов обычно делается предположение о том, что наблюдаемый поток отказов случаен, стационарен (однороден во времени), интенсивности отказов случайны, независимы и т.д. Однако часто эти предпосылки (исходные гипотезы) не выполняются. Следовательно, принятое окончательное решение о состоянии исследуемого объекта и все численные оценки надежности, принятые в данном случае, будут весьма сомнительны.

Под стационарностью (однородностью во времени) понимается тот факт, что интенсивности отказов будут одинаково распределены вне зависимости от того, где на временной оси находится интервал фиксированной длины. Случайность, в первую очередь, означает независимость наблюдаемых интенсивностей отказов, а также отсутствие в них каких-то простых закономерностей, трендов (тенденций). Наличием закономерности, к примеру, можно считать ситуацию, когда поток содержит временной промежуток (иногда не один) с существенно большей интенсивностью отказов, нежели в среднем, на всем наблюдаемом интервале времени. Простой линейный тренд говорит о наличии близкой к линейной зависимости между наблюдаемой частотой отказов и временем $t$. Нас будут интересовать статистические критерии случайности предназначенные для обнаружения линейных трендов.

Для авторов данной статьи такого рода критерии потребовались при проведении статистического анализа надежности систем управления защитой Билибинской АЭС (СУЗ БиАЭС). Исходные для расчетов данные представляли собой группированный поток отказов– количества (частоты, интенсивности) отказов за конкретный год от начала эксплуатации по настоящее время для каждого элемента СУЗ. На первоначальном этапе анализа необходимо было проверить гипотезу случайности этих

частот, в частности, отвергнуть гипотезу старения, выражающуюся в постепенном увеличении интенсивности отказов.

Некоторые методы решения подобных задач статистического анализа данных требуют усовершенствований, модернизаций, а иногда и разработки, особенно в том случае, когда исходная статистическая информация обладает определенного рода недостатками. К такого рода недостаткам можно отнести наличие в исходных данных повторяющихся наблюдений (связи в выборке). Так статистические данные об интенсивностях отказов СУЗ БиАЭС содержали в себе чрезвычайно много повторяющихся наблюдений, особенно нулей, поскольку отказы некоторых элементов были достаточно часто редкими событиями.

Для проверки случайности (отсутствия тренда) в этой ситуации достаточно часто применяется критерий Кендалла с поправками на связи П. Сена [1]. Но как известно, в ряде случаев, к примеру при малой выборке, критерий Кендалла менее мощный по сравнению с критерием Спирмена. Можно ожидать, что если корректно учесть поправки в критерии Спирмена, то применение нового критерия станет более предпочтительнее в смысле мощности. Кроме этого желательно провести сравнительный анализ мощности и интегрального критерия случайности, предложенного в работе [2].

В данной статье рассматривается методика проверки гипотезы о случайности статистических данных с учетом поправки на имеющиеся связи. При этом понимается, что принятая гипотеза о случайности, будет на самом деле означать отсутствие монотонных трендов в наблюдаемом временном ряде. Кроме этого методом статистических испытаний исследуются мощности различных критериев против альтернативы наличия монотонного тренда.

Сделаем краткое описание гипотезы случайности.

## Гипотеза случайности

Достаточно часто в качестве альтернативы гипотезе случайности выступает гипотеза о начале периода приближения системы к предельному состоянию. В этом случае интенсивности потока отказов имеют явную тенденцию к возрастанию. Если рассматриваются отказы новой системы, то в силу известного явления приработки, интенсивности потока отказов будут относительно высоки вначале исследуемого временного промежутка с постепенной тенденцией их уменьшения. В этом случае альтернативой нулевой гипотезе будет предположение о наличии отрицательного тренда в наблюдаемых частотах. Если исследователя интересует просто наличие монотонного тренда (без указания убывающий он или возрастающий), то рассматривают двустороннюю альтернативу.

*Гипотеза случайности* – это, пожалуй, первая и фундаментальная гипотеза, применяемая при обработке числовых случайных данных. Она заключается в предположении отсутствия тенденций и детерминированных закономерностей в этих данных. Сформулируем ее.

В различных статистических задачах исходные данные $\mathbf{X}=(X_1,..., X_n)$ часто рассматривают как случайную выборку из некоторого распределения $L(\xi)$, т.е. считают компоненты $X_i$ вектора данных $\mathbf{X}$ независимыми и одинаково распределенными случайными величинами. Как правило, это предположение оправдано и вытекает из самого характера задачи, но иногда оно нуждается в проверке.

Гипотеза случайности $H^*$ состоит в предположении симметричности функции (или плотности) распределения [3].

$$H_* : F_{\mathrm{X}}(x_{r_1}, x_{r_2}, ..., x_{r_n}) = F_{\mathrm{X}}(x_1, x_2, ..., x_n), \qquad (1)$$

Часто рассматривают упрощенный вариант этой гипотезы, дополнительно предполагая наличие независимости компонент.

$$H_0 : F_{\mathbf{X}}(x_1, ..., x_n) = F(x_1)...F(x_n), \qquad (2)$$

где $F(x)$– некоторая функция распределения. Такую гипотезу называют гипотезой случайности, хотя на самом деле в ней утверждается независимость и одинаковая распределенность компонент вектора $\mathbf{X}$. Таким образом, при $H_0$ рассматриваются независимые и одинаково распределенные случайные величины $X_1,X_2,...,X_n$. Очевидно, что $H_0 \subset H^*$, т.е. $H_0$– более жесткое предположение о характере исходных данных, нежели $H^*$. Тем не менее, именно эта нулевая гипотеза– $H_0$ будет проверяться в дальнейшем.

В непараметрической постановке целесообразно рассматривать следующие альтернативы [4].

Например:

$H_1^{(+)}$: $F_1(x) < F_2(x) < ... < F_n(x)$ – альтернатива возрастания,

$H_1^{(-)}$: $F_1(x) > F_2(x) > ... > F_n(x)$ – альтернатива убывания.

Критерий согласия для проверки гипотезы $H_0$ можно построить, исходя из различных соображений. Во-первых, предполагается, что вектор $\mathbf{X}$ имеет непрерывное распределение. Если гипотеза случайности действительно имеет место, то компоненты вектора $\mathbf{X}$ «равноправны» и поэтому данные не должны быть ни в каком смысле упорядочены. Другими словами, ситуацию, соответствующую гипотезе $H_0$, можно охарактеризовать как «полный хаос», или «полный беспорядок». При отклонениях от $H_0$ исходные данные имеют тот или иной порядок, проявляются связи, или зависимости от порядка индексации. Следовательно, критерий проверки $H_0$ можно построить на основании статистик, измеряющих степень «беспорядка» исходных данных.

## Связи в выборке

Этот раздел будет посвящен исследованию различных походов к учету связанных наблюдений. Как известно [1], *группой связанных наблюдений* называют множество наблюдений, имеющих одно и то же значение. «Когда исследователь ранжирует объекты на основе субъективных суждений, то это свойство (отсутствие предпочтений) связано с истинной их неразличимостью или неспособностью исследователя найти существующие различия. В этих случаях говорят, что такие объекты являются *связанными».* Не обязательно в этом случае говорить о том, что наблюдается дискретная случайная величина. На самом деле, связи могут появиться, к примеру в случае проведения округлений с заданной точностью непрерывной величины. При этом в исходной информации может наблюдаться несколько связанных групп. Если наблюдение имеет значение, отличающееся от остальных элементов выборки, то его можно считать группой связей объема 1.

В литературе обсуждались два метода обращения с совпадениями. Первый состоит в том, чтобы упорядочить совпавшие наблюдения случайным образом. Его достоинством является простота, он не требует новой теории, но при этом мы, очевидно, жертвуем информацией, содержащейся в наблюдениях, и можно ожидать, что он будет менее эффективен, чем второй метод, состоящий в том, что каждому из группы совпавших наблюдений приписывается средний ранг этой группы. Достоинства обоих методов исследовались довольно мало, но показано, что критерий Вилкоксона при случайном разделении совпадений имеет меньшую АОЭ, чем когда придаются средние ранги.

До появления дальнейшей информации метод среднего ранга кажется более общеупотребительным. По этой причине этот метод присваивания среднего ранга применяется и в этой статье.

## Свободный от распределения критерий независимости и случайности Кендалла

Известная статистика Кендалла $T_n$ может применяться для проверки гипотезы независимости одного набора данных (допустим $(X_1,...,X_n)$ от другого $(Y_1,...,Y_n)$ [5]). В этом случае исходной информацией является двумерный массив:

$$\begin{pmatrix} X_1 & X_2 & ... & X_n \\ Y_1 & Y_2 & ... & Y_n \end{pmatrix}. \qquad (3)$$

Столбцы этого массива переставляются таким образом, чтобы нижняя строка $Y$-ов была упорядочена (допустим по возрастанию). Если после этого и в верхней строке будет заметна тенденция к возрастанию (или убыванию) $X$-ов, то можно судить о наличии положительной (отрицательной) корреляционной зависимости $X$ от $Y$. Степень этой зависимости можно измерить числом инверсий $T_n$. Для такого рода данных статистика $T_n$ является оценкой величины $\tau$, которая определяется следующим образом:

$\tau = 2P\,((X_1\text{-}X_2)(Y_1\text{-}Y_2)>0)\text{-}1,$

где $P$– здесь и далее вероятность. Приведем критерий Кендалла в виде [1]. Однако заметим, что приведенная ниже статистика $K$ по сути является числом инверсий Кендалла $T_n$ для $X$-ов в том случае, когда столбцы двумерного массива (3) упорядочиваются в порядке возрастания $Y$-ов.

Кроме этого заметим, что если заменить $Y_i=i$ для каждого $i$, то критерий независимости превратится в критерий случайности в смысле предыдущего раздела. Альтернатива $\tau>0$ ($\tau<0$) будет означать наличие положительного (отрицательного) тренда у $X$-ов.

Для проверки гипотезы о независимости случайных величин $X$ и $Y$ (откуда следует $\tau=0$), а именно

$$H_0 : P\left(\{X \le a\} \cap \{Y \le b\}\right) = \\ = P(x \le a)\cdot P(Y \le b) \quad \forall a, b, \qquad (4)$$

надо проделать следующее.

Положить

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}\left(\left(X_i - X_j\right)\left(Y_i - Y_j\right)\right),$$

$$\text{где } \operatorname{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \qquad (5)$$

Таким образом каждой паре индексов $(i,j)$ приписывается $+1$, при $i<j$, если

$(X_i–X_j)(Y_i–Y_j) > 0$, и $–1$ иначе. Складывая эти единицы (с их знаками), мы получаем сумму $K$.

Известно, что при больших выборках следует использовать:

$$K^* = \frac{K - E_0(K)}{\sqrt{D_0(K)}} = \frac{K}{\sqrt{n(n-1)(2n+5)/18}}, \qquad (6)$$

где $E_0(K)$ и $D_0(K)$– математическое ожидание и дисперсия статистики $K$ при верной нулевой гипотезе случайности, $n$– объем выборки.

При выполнении $H_0$ статистика $K^*$ асимптотически (при $n\to\infty$) имеет стандартное нормальное распределение– $N(0,1)$.

Приближенный критерий в этом случае будет следующим:

$$\text{отклонить } H_0, \text{ если } K^* \ge u(\alpha), \qquad (7)$$
$$\text{принять } H_0, \text{ если } K^* < u(\alpha),$$

где константа $u(\alpha)$– квантиль стандартного нормального закона, т.е. величина, удовлетворяющая уравнению

$$P\,[K^* \le u(\alpha)] = \alpha.$$

Если среди $n$ наблюдений $X$ или среди $n$ наблюдений $Y$ есть связи, то дисперсию $D_0(K)$ в определении $K^*$ надо заменить на

$$D_0(K)=\frac{5}{12}\frac{\left(\begin{array}{l} n(n-1)(2n+5)- \\ -\sum_{i=1}^{g} t_i(t_i-1)(2t_i+5)- \\ -\sum_{j=1}^{h} u_j(u_j-1)(2u_j+5) \end{array}\right)}{18}+$$
$$+\frac{\left\{\sum_{i=1}^{g} t_i(t_i-1)(t_i-2)\right\}\left\{\sum_{j=1}^{h} u_j(u_j-1)(u_j-2)\right\}}{9n(n-1)(n-2)}+$$
$$+\frac{\left\{\sum_{i=1}^{g} t_i(t_i-1)\right\}\left\{\sum_{j=1}^{h} u_j(u_j-1)\right\}}{2n(n-1)} \qquad (8)$$

где $g$– число групп совпадающих наблюдений $X$; $t_i$– объем $i$-й группы наблюдений $X$; где $h$– число групп совпадающих наблюдений $Y$; $u_j$– объем $j$-й группы наблюдений $Y$. Поправка (8) впервые была получена П.Сеном.

## Ранговый критерий Спирмена

В своих работах для проверки гипотезы $H_0$ (1) Спирмен предложил следующую меру линейной связи между случайными величинами.

$$R=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(r_i^x-\frac{n+1}{2}\right)\left(r_i^y-\frac{n+1}{2}\right), \qquad (9)$$

где $r_i^x$ и $r_i^y$ – ранги $X_i$ и $Y_i$. Массивы упорядочиваются раздельно. $R$ называется *коэффициентом ранговой корреляции Спирмена* [6].

Поскольку от перемены мест слагаемых сумма не меняется, можно столбцы массива (3) упорядочить так, чтобы $Y$ возрастали и затем определить $r_i$– получившиеся (относительные) ранги $X_i$. В этом случае можно получить следующий вид ранговой статистики Спирмена [1]:

$$R=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)\left(r_i-\frac{n+1}{2}\right). \qquad (11)$$

Иногда пользуются и более удобной для расчета формой [5]:

$$\rho=1-\frac{6}{n(n^2-1)}\sum_{i=1}^{n}(i-r_i)^2. \qquad (12)$$

Покажем, что $R$=ρ *при отсутствии связей*:

$$R=\frac{12}{n^3-n}\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)\left(r_i-\frac{n+1}{2}\right)=$$
$$=\frac{12}{n^3-n}\left(\sum ir_i-n\frac{(n+1)^2}{4}\right).$$

$$\rho=1-\frac{6}{n^3-n}\sum_{i=1}^{n}(r_i-i)^2=\frac{12}{n^3-n}\left(\sum ir_i-n\frac{(n+1)^2}{4}\right).$$

*При наличии одной связной группы объемом t*:

$$R=\frac{12}{n^3-n}\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)\left(r_i-\frac{n+1}{2}\right)=$$
$$=\frac{12}{n^3-n}\left[\sum ir_i-n\frac{(n+1)^2}{4}\right]. \qquad (10)$$

$$\rho=1-\frac{6}{n^3-n}\sum_{i=1}^{n}(r_i-i)^2=$$
$$=\frac{12}{n^3-n}\left(\sum ir_i-n\frac{(n+1)^2}{4}+\frac{t^3-t}{24}\right). \qquad (11)$$

Из вышесказанного можно сделать вывод, что при отсутствии связей $R$=ρ, но при наличии связей коэффициентом ρ пользоваться не совсем правильно. Ошибка в первую очередь, будет связана с тем, что математическая форма статистики $R$ больше соответствует оценке ранговой корреляции, нежели удобная для расчетов статистика ρ.

Поэтому, взяв за основу статистику $R$ рассчитаем поправки на дисперсию, аналогичные поправкам П. Сена для статистики нормализованной статистики критерия Кендалла $K^*$(6).

## Дисперсия статистики Спирмена при наличии связных групп

При больших размерах выборки лучше использовать нормализованный критерий Спирмена, т.е. критерий, чья статистика при $H_0$ приближенно будет иметь стандартный нормальный закон.

То, что статистика Спирмена при верной $H_0$ имеет асимптотически нормальное распределение доказано, например, в монографии [1]. Стандартизации нормального закона можно добиться аналогично (6).

$$R^*=\frac{R-E_0(R)}{\sqrt{D_0(R)}}=\sqrt{n-1}\cdot R, \qquad (12)$$

т.к. математическое ожидание $E_0(R)$=0 и дисперсия $D_0(R)$= 1/($n$–1).

При выполнении $H_0$ статистика $R^*$ имеет асимптотическое (при $n\to\infty$) распределение $N(0,1)$.

Найдем дисперсию статистики критерия Спирмена при наличии связных групп.

Далее приведены расчеты дисперсии статистики Спирмена $R$ без учета коэффициента $12/(n^3-n)$, т.е. статистики $\tilde{R}=\frac{R(n^3-n)}{12}$.

Обозначим $c_i=i-\frac{n+1}{2}$. Предположим, что имеется одна связанная группа объема $t$.

$$D_0\left(\tilde{R}\right)= D_0\left[\sum_{i=1}^{n} c_i r_i\right]=$$
$$= \sum_{i=1}^{n} c_i^2 D_0\left(r_i\right)+ 2\sum_{i<j} c_i c_j cov_0\left(r_i, r_j\right)=$$
$$= \sum c_i^2 \left(D_0\left(r_i\right)- cov_0\left(r_i, r_j\right)\right)$$

Определим дисперсию ранга $D_0\left(r_i\right)$ при верной нулевой гипотезе:]

$$D_0\left(r_i\right)= E_0\left(r_i^2\right)- E_0^2\left(r_i\right)= \frac{n^2-1}{12}- \frac{t(t^2-1)}{12(n-1)},$$

где $E_0$ – математическое ожидание.

Ковариация рангов $r_i$ и $r_j$:

$$cov_0(r_i, r_j) = E_0\left(r_i r_j\right)- E_0\left(r_i\right)E_0\left(r_j\right)=$$
$$= \frac{n^2(n+1)^2}{12}- \frac{n(n+0.5)(n+1)}{3n(n+1)}- \frac{(n+1)^2}{4}.$$

Т.к. $\sum c_i^2 = \frac{n(n-1)(n+1)}{12}$, то после некоторых несложных вычислений получаем

$$D_0\left(\tilde{R}\right)= \frac{n^2(n+1)^2(n-1)}{144}\left(1- \frac{t(t^2-1)}{n(n^2-1)}\right).$$

Таким образом, если число связных групп равно $k$, то

$$D_0\left(\tilde{R}\right)= \frac{n^2(n+1)^2(n-1)}{144}\left(1- \frac{\sum_{i=1}^{k} t_i(t_i^2-1)}{n(n^2-1)}\right).$$

Окончательная форма нормализованной статистики критерия Спирмена имеет вид:

$$R^* = \frac{\tilde{R}}{\sqrt{D_0(\tilde{R})}} = \frac{\sum_{i=1}^{n}\left(i- \frac{n+1}{2}\right)\left(r_i- \frac{n+1}{2}\right)}{\sqrt{\frac{n^2(n+1)^2(n-1)}{144}\left(1- \frac{\sum_{i=1}^{k} t_i(t_i^2-1)}{n(n^2-1)}\right)}}. \quad (13)$$

где $k$– количество связанных групп, $t_i$– объем $i$-й связанной группы.

При отсутствии связей получаем (13). Таким образом приближенный критерий Спирмена при наличии связей будет следующим:

отклонить $H_0$, если $R^* \geq u(\alpha)$, (14)
принять $H_0$, если $R^* < u(\alpha)$,

где константа $u(\alpha)$– квантиль стандартного нормального закона, т.е. величина, удовлетворяющая уравнению $P\,[R^* \leq u(\alpha)]=\alpha$.

## Интегральный критерий

В работе [2] был предложен интегральный критерий, построенный на статистиках следующего вида:

$$In^{(++)} = \frac{\sum_{i=1}^{n}\left(\sum_{1}^{i} r_j - \overline{r^{(+)}}\right)\left(\sum_{1}^{r_i} j - \overline{j^{(+)}}\right)}{S_r^{(+)} S_j^{(+)}};$$

$$In^{(-+)} = \frac{\sum_{i=1}^{n}\left(\sum_{i}^{n} r_j - \overline{r^{(-)}}\right)\left(\sum_{1}^{r_i} j - \overline{j^{(+)}}\right)}{S_r^{(-)} S_j^{(+)}}; \quad (15)$$

где $\overline{r^{(+)}} = \frac{1}{n}\sum_{i=1}^{n}(n+1-i)r_i$; $\overline{r^{(-)}} = \frac{1}{n}\sum_{i=1}^{n} i r_i$;

$$S_r^{(+)} = \sqrt{\sum_{i=1}^{n}\left(\sum_{j=1}^{i} r_j - \overline{r^{(+)}}\right)^2}; \; S_r^{(-)} = \sqrt{\sum_{i=1}^{n}\left(\sum_{j=i}^{n} r_j - \overline{r^{(-)}}\right)^2};$$

$$\overline{j^{(+)}} = \frac{(n+1)(n+2)}{6}- \frac{1}{24n}\sum_{l=1}^{s} t_l\left(t_l^2-1\right);$$

$$S_j^{(+)} = \sqrt{\sum_{i=1}^{n}\left(\frac{r_i\left(r_i+1\right)}{2}- \overline{j^{(+)}}\right)^2}.$$

Во избежание асимметрии распределения предлагается в качестве конечного варианта статистики использовать линейную комбинацию исходных (15), например:

$$In^{(\circ+)} = 0,5(In^{(++)} - In^{(-+)}), \quad (16)$$

В этом случае любая статистика будет несмещенной при условии выполнения нулевой гипотезы. Для статистики вида (19) построено табличное обеспечение и доказано, что в асимптотике данная статистика имеет то же распределение, что и статистика Спирмена [2]. Для объема наблюдений $n>7$ получены критические значения с помощью улучшенного нормального приближения для распределения статистики Спирмена [1]:

$$In_\alpha = \frac{u(\alpha)}{\sqrt{n-1}}\left\{1- \frac{0.19}{n-1}\left[u^2(\alpha)-3\right]\right\},$$

где константа $u(\alpha)$– квантиль стандартного нормального закона.

Приближенный интегральный критерий при наличии связей будет следующим:

отклонить $H_0$, если $In^{(\circ+)} \geq In_\alpha$, (17)
принять $H_0$, если $In^{(\circ+)} < In_\alpha$,

Дальнейшие исследования будут касаться сравнительного анализа критериев Кендалла, Спирмена и интегрального критерия при наличии связей.

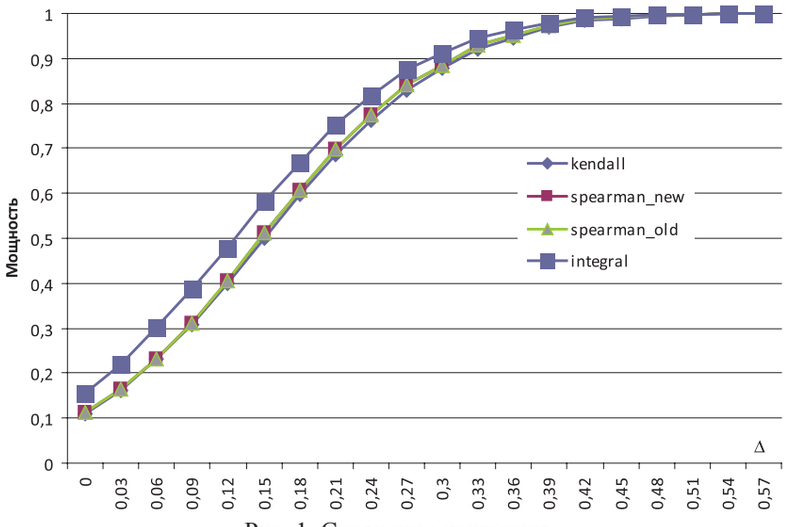


Рис. 1. Сравнение критериев

## Сравнение критериев

Как известно, одним из основных методов сравнения статистических критериев является анализ их мощности– условной вероятности принять альтернативу при условии, что она верна. Функция мощности для каждого критерия зависит от выбранного уровня значимости $\alpha$ и некоторого параметра $\Delta$ (или вектора параметров $\vec{\Delta}$ и от объема данных). В нашем случае алгоритм сравнения критериев проверки гипотезы случайности по мощности был основан на обычном методе Монте-Карло и построен следующим образом.

Многократно моделируется гауссовская (нормальная) псевдослучайная выборка $X_1, X_2, ..., X_n$ фиксированного объема $n$ с заданной структурой связей и с постепенно увеличивающимся положительным трендом $\Delta$ по следующей формуле:

$$X_i=\xi_i+i\cdot\Delta,$$

Моделируются связи заданных объемов. Процесс генерации выборки повторяется $m$ раз. Мощность $P_1(\Delta)$ оценивается эффективной оценкой

$$P_i(\Delta)=\frac{\nu_i(\Delta)}{m}.$$

Далее увеличивается угол тренда. Повторяются шаги, связанные с генерацией выборки $m$ раз и подсчетом функции мощности.

На рис. 1 приведены графики мощности четырех критериев:

1. критерия Кендалла с поправкой на связь Сена (7), (8) – Kendall;
2. критерия Спирмена, построенного на статистике без учета связей (12)– Sperman_old;
3. критерия Спирмена (14) с поправкой на связь (13)– Sperman_new;
4. интегрального критерия (17).

Объемы выборок $n$=10.

Как видно из графика, при малых размерах выборки интегральный критерий обладает большей мощностью, но имеет небольшое смещение, вызванное ошибкой нормального приближения. При увеличении объема выборки естественно ожидать уменьшения этого смещения.

## Заключение

В статье рассмотрен вопрос проверки гипотезы случайности (отсутствия линейного тренда) различными статистическими критериями при наличии в исходных данных связей– одинаковых значений.

В частности решены следующие задачи:

Разработан нормализованный критерий Спирмена с поправкой на связь.

Проведены эксперименты по сравнению мощности критериев. В рамках этих экспериментов выявлено некоторое преимущество интегрального критерия перед остальными.

## Литература

1. **Холлендер М., Вулф Д.** Непараметрические методы статистики.-М.: – Финансы и статистика, 1983.- 518 с

2. **Антонов А.В., Чепурко В.А.** Интегральный критерий проверки гипотезы тренда. Надежность.-2006.-№ 2.- с.17-27

3. **Гаек Я., Шидак З.** Теория ранговых критериев. – М.: – Наука, 1971.-376с

4. **Буртаев Ю.Ф., Острейковский В.А.** Статистический анализ надежности объектов по ограниченной информации. –М.: Энергоатомиздат, 1995, -240с

5. **Кендэл М.** Ранговые корреляции.- Зарубежные статистические исследования. М., «Статистика», 1975.-216 с.

6. **Ван дер Варден** Математическая статистика М.: Изд. Иностр. лит., 1960.– 435 с.