УДК 621.398

© Казаков Г.В.
Kasakov G.

КРИТЕРИИ АНАЛИЗА СВЯЗАННЫХ ВЫБОРОК ПРИ ИСПЫТАНИЯХ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ АСУ

CRITERIA ANALYSIS RELATED SAMPLES IN TESTING SOFTWARE ACS

Аннотация. *В зависимости от условий испытаний рассмотрены параметрические и непараметрические критерии принятия решений о близости связанных выборок при испытаниях программных средств АСУ и принятие их в эксплуатацию.*

Annotation. *Depending on the test conditions considered parametric and nonparametric decision-making related to the proximity of samples when testing software ACS and their acceptance into operation.*

Ключевые слова. *Испытание, программное обеспечение АСУ, связанная выборка, поддержка решений, параметрический критерий, непараметрический критерий.*

Key words. *Испытание, программное обеспечение АСУ, связанная выборка, поддержка решений, параметрический критерий, непараметрический критерий.*

В практике испытаний и ввода в эксплуатацию специального программного и математического обеспечения АСУ широко реализуется принцип совместных испытаний различных версий программных средств, предназначенных для решения однотипных задач. Совместные испытания программного обеспечения (ПО) АСУ проводятся с целью повышения уверенности в его функциональной готовности и подтверждения характеристик его качества: оценки близости функций распределения, оценки близости математических ожиданий и дисперсий выборок, оценки допусков на характеристики точности, в пределах которых можно уверенно говорить о пригодности ПО к применению. Эти оценки могут быть получены с использованием связанных выборок.

Связанными случайными выборками обычно называются такие, которые получены в экспериментах, проведенных с объектами измерений в одних и тех же (или близких) условиях. При испытаниях ПО сравниваются модели, разработанные на базе так называемых унаследованных компонентов, когда осуществляются измерения одних и тех же характеристик ПО, а отличие в их оценках обусловлено разработкой моделей в различных организациях, некоторой модификацией алгоритмов, применением различных вычислительных средств, алгоритмов различной сложности, различных языков программирования, различных операционных систем и т.п. При этом возникает вопрос о порядке проверки достоверности, целостности, точности и устойчивости функционирования такого ПО. Данное исследование посвящено вопросам контроля точности характеристик ПО.

1. Параметрические критерии проверки близости связанных выборок

1.1. *Критерии близости, основанные на оценке среднего*

Обычно одну из связанных выборок принимают за эталонную и считают разность связанных выборок, распределенной по нормальному закону. Тогда степень близости выборок может быть легко определена методом доверительных (с заданным уровнем доверия) интервалов для среднего с использованием критерия Стьюдента (для малых выборок) или нормального распределения (для больших выборок) [1,2].

В специальных задачах, когда требуется, чтобы попадание в интервал с доверительной вероятностью (1-a) составляло заданную долю p значений контролируемых выборок, используют толерантные интервалы [3]. В работе [3] же приведены таблицы для объемов испытаний в функции доверительной вероятности и доли p значений контролируемых выборок.

Однако в этом случае имеют место отдельные, не всегда оправданные допущения. Таким допущением является, в частности, принятие одной из связанных выборок за эталонную для связанных выборок одинаково-

Казаков Геннадий Викторович – кандидат технических наук, доцент, начальник управления, ФБУ «4 ЦНИИ МО РФ», тел.8 (910) 467- 71- 66.

Kasakov Gennadiy – candidate of technical science, bead of department, FBU «4 CSRI MD RF», tel. (910) 467-71-66.

го или близкого уровня точности. Для правильного назначения допусков возможного расхождения связанных выборок необходимо всегда уточнять закон распределения как для исходных выборок, так и их разностей. Если при этом отсутствуют какие-либо теоретические предпосылки, необходимо проверить гипотезу о виде закона распределения разности двух случайных связанных выборок. Существует множество методик и критериев для проверки соответствия распределения эмпирической выборки наперед заданному теоретическому распределению, особенно для больших выборок.

Для проверки принадлежности к нормальному распределению (в теории ошибок отдается предпочтение именно ему) для малых выборок, объемом от 3 до 50 вариантов, следует использовать наиболее мощный для проверки гипотезы нормальности критерий Шапиро-Уилка [5]

$$U = \frac{b^2}{s^2};\; s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2;\; b = \sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i),$$

где x_i, $(i = 1,...,n)$ – ранжированный ряд;

$k=n/2$, если n – четное;

$k=(n-1)/2$, если n – нечетное;

a_{n-i+1}, $(i = 1,...,k)$ – константы.

Таблица распределения статистики критерия приведена в работе [2].

Например, для ранжированной выборки: 303, 338, 406, 457, 461, 469, 474, 489, 515, 583, уровня значимости – 50% (критическое значение статистики – 0,938211), рассчитанное значение статистики – 0,9433039. Это значение превышает критический уровень. Нет оснований предполагать, что представленная выборка не подчинена нормальному закону распределения.

Однако следует иметь в виду, что допуски нормального закона распределения без учета вида реального закона распределения бывают существенно занижены. Установлено [4], что границы доверительных интервалов допусков для распределений Лапласа, логистического, максимальных и минимальных значений шире, чем при нормальном распределении. Верхняя граница допустимых значений в этих случаях для доверительной вероятности 0,997 составит 4,1s, 3,6s, 4s, 4s соответственно.

Если заранее известны или определены распределения для каждой из независимых исходных выборок, то, пользуясь правилом декомпозиции для суммы случайных величин вида $x = x_1 + (-x_2)$, можно записать

$$f(x) = \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(x_1 - x)dx_1.$$

Следует помнить, что при одних и тех же распределениях слагаемых распределение разностей может быть разным в зависимости от совместного распределения слагаемых

$$f(x) = \int_{-\infty}^{\infty} f_1(x_1 - x)dx_1.$$

Так, если распределения каждой из независимых исходных выборок подчинены нормальному закону распределения, т.е. $x_1 \in N(m_1, \sigma_1)$, $x_2 \in N(m_2, \sigma_2)$ и

$$p(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp(-\frac{(x_1 - m_1)^2}{2\sigma_1^2});$$

$$p(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \cdot \exp(-\frac{(x_2 - m_2)^2}{2\sigma_2^2}),$$

то выражение для плотности вероятности случайной величины x с учетом (1) будет иметь вид

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp(-\frac{(x_1 - m_1)^2}{2\sigma_1^2}) \times$$

$$\times \frac{1}{\sqrt{2\pi}\sigma_2} \cdot \exp(-\frac{(x_2 - m_2)^2}{2\sigma_2^2})dx_1.$$

После соответствующих преобразований (можно с применением символьной математики) выражение для плотности вероятности разностей связанных выборок примет вид

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \cdot \exp\left[-\frac{(x - (m_1 - m_2))^2}{2(\sigma_1^2 + \sigma_2^2)}\right]$$

и это будет плотность нормального закона распределения разности связанных выборок с параметрами

$m = m_1 + m_2$ и $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$.

На графиках (рис.1) приведены функции распределения x связанных выборок и функция распределения и плотность вероятности их разностей.

Это служит еще одним подтверждением устойчивости нормального закона распределения при его различных композициях.

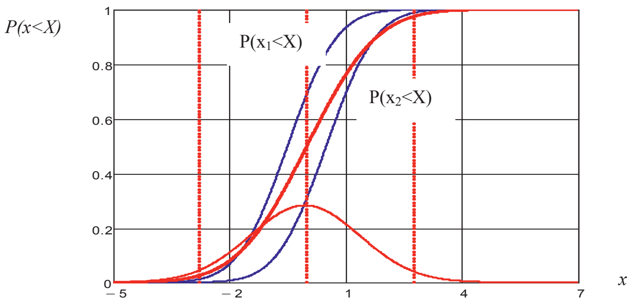

Рис.1. Функции распределения связанных выборок и функция распределения и плотность вероятности их разностей

1.2. *Оценка допусков на параметры с использованием функций распределения размаха невязок связанных выборок измерений*

Максимальное значение разности Δr – разности между наибольшим и наименьшим значениями случайной величины Δr называется размахом разностей двух связанных выборок [4].

Функция распределения размаха Δr

$$W = \max_{1<i<n} X_i \quad - \min_{1<i<n} Y_i$$

находится [6] как интеграл Стильтьеса

$$P(\Delta r < y) = n \cdot \int^{\infty} \left[F(x+y) - F(x)\right]^{n-1} dF(x).$$

Для нормального распределения размаха Δr она запишется

$$P(\Delta r < y) = n \cdot \int_{-\infty}^{\infty} \left[pnorm\left(x+y,m,\sigma\right) - \right.$$

$$\left. - pnorm\left(x,m,\sigma\right)\right] dnorm\left(x,m,\sigma\right) dx$$

или

$$P(\Delta r < y) = n \cdot \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x+y-m)^2}{2\sigma^2}\right) - \right.$$

$$\left. - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \right]^{n-1} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

На рис.2 представлены функция распределения разностей выборок и функции распределения размаха этой разности. Функция распределения случайной величины является исчерпывающей ее характеристикой, поэтому на ее основе можно легко определить как уровни доверительной вероятности, так и квантили ее распределения.

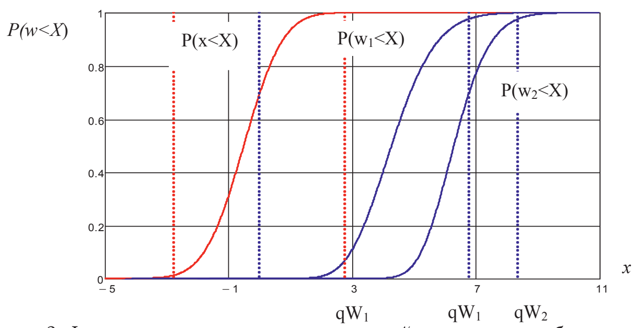

рис.2. Функция распределения разностей связанных выборок и их размахов для различных объемов выборок (n_1 и n_2)

Из рис. 2 следует, что принятие допусков для среднего Δr сужает их границы, как это имело бы место при принятии одной из выборок за эталонную, что необоснованно увеличивает требования к уровням близости связанных выборок. В то же время распределения реального размаха связанных выборок дают основания принять расширенные допуски, которые более объективно отражают процесс проверки их близости без риска получения недопустимого ущерба. То есть, если размах статистической выборки меньше квантили ее распределения для заданного уровня доверия β $W \le q_\beta$, то гипотеза о

близости связанных выборок верна.

Из рис. 2 также следует, что функция распределения размаха существенно зависит от объемов связанных выборок *n*. Прообразом статистического решения о выборе подходящей функции распределения может служить число применений ПО в практике его эксплуатации в АСУ.

В работе [6] также показано, что величина размаха W_n случайной выборки объема *n* позволяет определить ее СКО, в нашем случае – σ разностей связанных выборок Δr как $\sigma = M(W_n)/W_n$, где $M(W_n)$ – математическое ожидание размаха выборок для заданного их объема *n*.

1.3. *Оценка точности моделей ПО АСУ, разработанных в различных организациях*

Иногда бывает недостаточно ограничиться сравнением программных средств, разработанных в различных организациях, по так называемым «критериям правильности – допускам», но можно использовать имеющуюся информацию и для оценки погрешностей контролируемых параметров *q*.

При сравнении двух независимых выборок характеристик ПО, принадлежащих одному классу точности, можно определить выборочные дисперсии погрешностей испытываемых моделей

$$\sigma_q^2 = \frac{1}{\sqrt{2}} \sigma_{\Delta q}^2,$$

где $\sigma_{\Delta q}^2 = \dfrac{\sum_{i=1}^{n} (\Delta q_i - \overline{\Delta q})^2}{2(n-1)}$; $\overline{\Delta q} = \dfrac{\sum_{i=1}^{n} \Delta q_i}{n}$; $\Delta q_i = q_{1i} - q_{2i}$.

Проверка согласованности моделей (значимости величины $\overline{\Delta \mathrm{q}}$) осуществляется с помощью распределения Стьюдента, которому подчинена величина

$$t = \frac{\overline{\Delta q}\sqrt{n}}{\sigma_q \sqrt{2}}.$$

При наличии полностью независимых *k* тестовых выборок (в том числе по их объемам (N_1, ..., N_k) искомые оценки находятся как средневзвешенные из составных распределений

$$\sigma_q^2 = \sum_{j=1}^{k} P_j (\sigma_j^2 + (\overline{\Delta q}_j - \Delta q)^2),$$

где $\Delta \overline{q} = \sum_{j=1}^{k} P_j \Delta q_j$; $P_j = \frac{N_j}{\sigma_j^2}(q_{1i} - q_{2i})$;

$\Delta \overline{q}_j, \sigma_j^2$ – соответственно выборочное среднее и среднее квадратическое отклонение погрешностей j-совокупности.

Наконец, можно показать, что критерий Стьюдента для проверки гипотезы $H_0 : \Delta \overline{q} = 0$ против альтернативы $H_1 : \Delta \overline{q} \neq 0$ записывается в виде

$$t = \frac{\Delta \overline{q}(\sum_{i=1}^{k} N_i - k)}{\sqrt{\sum_{j=1}^{k}(N_j - 1)\sigma_{qj}^2 \sqrt{\frac{\sum_{i=1}^{k} N_i}{\prod_{i=1}^{k} N_i}}}}.$$

В работе [7] осуществлена проверка гипотез об однородности связанных выборок путем проверки гипотез о равенства медиан, математических ожиданий и совпадения функций распределения связанных выборок.

3. Непараметрические критерии проверки гипотез о различимости связанных выборок

Когда предпосылки применения параметрических критериев не выполняются, важно рассмотреть критерии, применимые в широком диапазоне условий. Методы, их использующие и применяемые в широком диапазоне условий, называются робастными [4] (понятие самой робастности как раз увязывается со слабой чувствительностью к отклонениям от стандартных условий). В этих условиях применяются непараметрические критерии, которые не требуют предварительных предположений относительно вида исходного распределения и являются более робастными, чем их параметрические аналоги. Тем более что на практике с ограниченными условиями испытаний мы сталкиваемся достаточно часто.

3.1. *Ранговый критерий оценки близости математических ожиданий связанных выборок*

Рассмотрим в этой связи ранговый *W*-критерий Вилкоксона [5], который может использоваться для решения различных задач: оценки близости функций распределения, оценки близости математических ожиданий для произвольных выборок и др.

$$T = \min\left(\sum_{i=1}^{n_1} R_i, \sum_{i=1}^{n_2} S_i\right),$$

где R_i, S_i – ранги (т.е. места в упорядоченных рядах) выборки, имеющей наименьшую и наибольшую сумму рангов.

С его помощью можно также выявить значимость различий в двух равных (n_1=n_2=N) или и неравных связанных выборках, отличных лишь воздействием на одну из них некоторого возмущающего фактора. Для практического вычисления критерия в этой задаче используется статистика [5]

$$\frac{W - E[W]}{\sqrt{D[W]}},$$

где $E[W] = N \cdot (N-1)/4$ – математическое ожидание;

$D[W] = N \cdot (N-1)(2N+1)/24$ – дисперсия;

N – величина выборки в каждом ряду, распределенная по стандартному нормальному закону.

По полученным двум выборкам R_{1i} и R_{2i}, (i=1,2, ..., N) находятся их разности $dR_i = R_{1i} - R_{2i}$, (i=1,2, ..., N) которые затем ранжируются.

Критерий проверяет статистическую значимость нулевой гипотезы H_0 о том, что распределение случайной величины $dR_i = R_{1i} - R_{2i}$, (i=1,2, ..., N) симметрично.

Рассмотрим пример. Пусть будут заданы две выборки, полученные на однотипном в основном составе исходных данных (табл. 1).

Таблица 1

Выборки исходных данных

№ п/п	1	2	3	4	5	6	7	8	9	10
Выборка 1	303	338	406	457	461	469	474	489	515	583
Выборка 2	305	341	411	452	459	460	472	481	511	580

За уровень значимости критерия примем величину 0,1. Требуется оценить значимость новых факторов неопределенности.

Рассчитанное значение статистики будет равно 0,050965. Следовательно, различия в выборках следует признать значимыми на уровне значимости 0,1 [1]. И, таким образом, влияние мешающих параметров, при принятии решений необходимо учитывать, т.е. идентифицировать их и учесть в моделях ПО. Формула и таблица распределения статистики критерия приведены в работе[1].

3.2. *Проверка гипотезы о равенстве дисперсий двух связанных выборок по критерию Ансари-Брэдли*

Еще одна модель, связанная с использованием другого важного рангового критерия - критерия Ансари-Брэдли [5], служит для статистической проверки нулевой гипотезы о равенстве дисперсий двух совокупностей. Ее также можно обоснованно применить как для объединения различных совокупностей испытаний в единую генеральную совокупность, так и для выявления значимости случайных факторов при их существенном расхождении. Вычисления статистики критерия производятся по следующим формулам:

$$W = \begin{cases} \left[\sum_{i=1}^{n_1} R_i - \frac{n+2}{4} \right] \sqrt{\frac{48 n_1 (n-1)}{n_2 (n-2)(n+2)}}, \\ \text{если } n \text{ четное;} \\ \left[\sum_{i=1}^{n_1} R_i - \frac{(n+1)^2}{4n} \right] \sqrt{\frac{48 (n-1)}{n_2 (n+1)(n+3)}}, \\ \text{если } n \text{ нечетное,} \end{cases}$$

где R_i, (i=1, ...,n_1) – ранги первой выборки;

n_1, n_2, n – объем первой, второй и в совокупности обеих выборок соответственно.

При существенном различии характеристик выборок на заданном уровне значимости делается вывод о том, что влияние мешающих параметров значимо. Их необходимо учитывать при применении программного изделия и постановке экспериментов по оценке характеристик точности и надежности программы. При существенном расхождении параметров рассматриваемые выборки нельзя простым образом объединить в единую генеральную совокупность.

Рассмотрим пример. Пусть будут заданы две выборки, полученные на однотипном в основном составе исходных данных (табл.2).

Таблица 2

Выборки исходных данных

№ п/п	1	2	3	4	5	6	7	8	9	10
Выборка 1	3,60	3,91	3,75	4,03	3,96	3,4	3,67	3,72	4,01	3,64
Выборка 2	3,65	3,84	3,77	3,82	3,76	3,54	3,80	3,73	3,76	3,69

За уровень значимости критерия примем величину 0,05 (уровень доверительной вероятности 0,95). Требуется оценить уровень значимости возможно новых появившихся факторов неопределенности случайного происхождения.

Рассчитанное значение статистики будет равно 0,050965. Следовательно, различия в выборках следует признать значимыми на уровне значимости 0,05 [1]. И, таким образом, влияние вновь возникших мешающих параметров значимо и их также необходимо учитывать при принятии решений о доработке второй модели с учетом неучтенных факторов или о расширении границ допусков при использовании второй модели.

Вместе с тем, всегда, когда предпосылки применения параметрических критериев выполняются, предпочтение следует отдавать им с тем, чтобы не допустить принятия ложной нулевой гипотезы (ошибки II рода).

Проверку на выполнение сходства качественных (логических переменных) параметров ПО легко осуществить в схеме испытаний Бернулли; при этом сходство связанных выборок по этим параметрам может быть обеспечено только при полном их совпадении.

Таким образом, в статье комплексно рассмотрены критерии анализа связанных выборок, получаемых при испытаниях ПО АСУ с использованием как параметрических, так и непараметрических критериев.

Литература
1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. - 416с.
2. Кендал М.Дж., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. - 899с.
3. ГОСТ Р 50779.22-2005. Статистические методы. Статистическое представление данных. Точечные оценки и доверительные интервалы для среднего.
4. Лесных Н.Б. Законы распределения случайных величин в геодезии: Монография. – Новосибирск: СГГА, 2005. –129 с.
5.мАнализ и обработка данных. Спецвыпуск. С-П:, М:, Х:, Мн:, Питер. - 2001. –751с.
6. Математический энциклопедический словарь. М.: Советская энциклопедия, 1988. - 845с.
7. Орлов А.И. Методы проверки однородности связанных выборок. //Заводская лаборатория, №7. 2004.

Материал поступил в редакцию 19. 03. 2014 г.