УДК 519.254

Критерии равенства математических ожиданий Стьюдента и Крамера–Уэлча для данных типа времени жизни

П. А. Филоненко[1], С. Н. Постовалов

Рассмотрены параметрический критерий Стьюдента и непараметрический критерий Крамера–Уэлча для проверки гипотезы однородности математических ожиданий. Данные критерии могут использоваться для проверки гипотезы однородности распределений. Представлены статистики критериев, применимые в случае данных типа времени жизни. Показано, что с ростом объемов выборок распределения статистик критериев приближаются к соответствующим предельным распределениям. Кроме того, в данной работе приведены результаты компьютерного моделирования мощности данных критериев с использованием имитационного метода Монте-Карло на альтернативных гипотезах при выполнении и нарушении априорных предположений.

Ключевые слова: проверка гипотезы однородности, критерий Стьюдента, критерий Крамера–Уэлча, параметрические критерии, непараметрические критерии, мощность критерия, распределение статистик, данные типа времени жизни, метод Монте-Карло.

1. Введение

В качестве проверки гипотезы однородности распределений могут рассматриваться критерии равенства математических ожиданий, критерии равенства дисперсий и их возможные комбинации в силу того факта, что однородность распределений означает совпадение всех характеристик распределений: $H_0 : S_1(t) = S_2(t)$ [1].

Если гипотеза однородности математических ожиданий не отклоняется, это еще не означает, что распределения $S_1(t)$ и $S_2(t)$ однородны, т.к. в таком случае могут отличаться любые другие характеристики распределения (например, дисперсии). Однако факт отклонения такой гипотезы может поставить под сомнение однородность распределений $S_1(t)$ и $S_2(t)$, т.к. как минимум одна характеристика распределений может различаться.

В математической статистике широкое распространение получил критерий Стьюдента для проверки однородности математических ожиданий [2]. Данный критерий относится к типу параметрических критериев, а используемым априорным предположением является тот факт, что данные распределены по гауссовскому (нормальному) закону распределения. Имеются работы, в которых показано, что при большом и близком объеме выборок n_1, n_2 нарушение априорного предположения на распределения статистик критерия Стьюдента S_t при справедливости нулевой гипотезы $G(S_t \mid H_0)$ существенного влияния не оказывает в си-

[1] Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках проектной части государственного задания (проект 2.541.2014К).

лу центральной предельной теоремы [3], однако в противном случае распределение $G(S_t \mid H_0)$ может существенно отличаться. Кроме того, проверка гипотезы о нормальности – более сложная статистическая процедура, чем проверка гипотезы однородности [4]. Поэтому в силу указанных обстоятельств предлагается [4] вместо описанного параметрического критерия Стьюдента использовать непараметрический критерий Крамера–Уэлча со статистикой S_{KW}. Из многомерной центральной предельной теоремы и из теорем о наследовании сходимости [5] следует, что с ростом объемов выборок n_1, n_2 распределение статистики Крамера–Уэлча $G(S_{KW} \mid H_0)$ сходится к предельному распределению (стандартный нормальный закон) и не имеет аналогичного априорного предположения о нормальности данных, как у критерия Стьюдента.

На практике далеко не всегда возможно произвести наблюдение за объектом от самого начала до самого конца исследования. Могут возникать ситуации, в которых объект выбывает из наблюдения, но наверняка известно, что до определенного момента времени этот объект был исправен. Такие наблюдения будем называть цензурированными [6], и для таких данных описанные выше критерии не могут быть применены.

Учитывая все вышесказанное, в главе 1 рассмотрим статистики критериев Стьюдента и Крамера–Уэлча, а также модификации этих критериев для случая данных типа времени жизни. Кроме того, с использованием метода Монте-Карло продемонстрируем, что распределение статистик модифицированных критериев $G(\hat{S}_t \mid H_0)$ и $G(\hat{S}_{KW} \mid H_0)$ с ростом объемов выборок n_1, n_2 приближается к соответствующим предельным распределениям. В главе 2 представим результаты компьютерного моделирования мощности критериев на альтернативных гипотезах при выполнении и нарушении априорных предположений о нормальном распределении данных.

2. Статистические методы

2.1. Постановка задачи

Предположим, мы имеем две выборки из непрерывных случайных величин ξ_1 и ξ_2:

$$X_1 = \{t_{11}, \ldots, t_{1n_1}\}, X_2 = \{t_{21}, \ldots, t_{2n_2}\}.$$

Выборкам X_1 и X_2 соответствуют функции надежности $S_i(t)$ с математическими ожиданиями μ_i и дисперсиями $\sigma_i^2, i = 1, 2$. Тогда

$$t_{ij} = \left(Z_{ij}, \delta_{ij}\right),$$

где $Z_{ij} = \min\left(T_{ij}, C_{ij}\right)$, $\delta_{ij} = \begin{cases} 1, T_{ij} \le C_{ij}, \\ 0, T_{ij} > C_{ij}. \end{cases}$, T_{ij} – моменты времени отказа, C_{ij} – моменты времени цензурирования для *j*-ого объекта в *i*-й группе. Величины T_{ij} и C_{ij} являются независимыми и одинаково распределенными с соответствующими функциями распределения $F_i(t)$ и $F_i^C(t)$. Функцией надежности $S(t)$ будем называть такую функцию, значение которой есть вероятность события, что не произойдет отказ объекта после времени t: $S_i(t) = P\{\xi_i > t\} = 1 - F_i(t)$.

В таком случае нулевая гипотеза H_0 об однородности математических ожиданий выглядит следующим образом:

$$H_0 : \mu_1 = \mu_2.$$

2.2. Классический критерий Стьюдента для полных наблюдений

В случае классического критерия Стьюдента для полных наблюдений будем считать $\delta_{ij}=1, \forall i,j$. В таком случае статистика критерия S_t будет выглядеть следующим образом:

$$S_t=\frac{\hat{\mu}_1-\hat{\mu}_2}{\sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2}}\sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}},$$

где

$$\hat{\mu}_1=\frac{1}{n_1}\sum_{i=1}^{n_1} Z_{1i},\ \hat{\mu}_2=\frac{1}{n_2}\sum_{i=1}^{n_2} Z_{2i},$$

$$s_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(Z_{1i}-\hat{\mu}_1)^2,\ s_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Z_{2i}-\hat{\mu}_2)^2.$$

Основная гипотеза H_0 отвергается с уровнем значимости α, если $|S_t|>t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$, где $t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$ – это $\left(1-\frac{\alpha}{2}\right)$квантиль распределения Стьюдента со степенью свободы n_1+n_2-2.

2.3. Классический критерий Крамера–Уэлча для полных наблюдений

В случае классического критерия Крамера–Уэлча [7] для полных наблюдений будем считать $\delta_{ij}=1, \forall i,j$. В таком случае статистика критерия S_{KW} будет выглядеть следующим образом:

$$S_{KW}=\frac{\sqrt{n_1 n_2}\,(\hat{\mu}_1-\hat{\mu}_2)}{\sqrt{n_2 s_1^2+n_1 s_2^2}}.$$

Основная гипотеза H_0 отвергается с уровнем значимости α, если $|S_{KW}|>z_{\left(1-\frac{\alpha}{2}\right)}$, где $z_{\left(1-\frac{\alpha}{2}\right)}$ – это $\left(1-\frac{\alpha}{2}\right)$ квантиль стандартного нормального распределения.

2.4. Оценки математического ожидания и дисперсии для данных типа времени жизни

Учитывая тот факт, что оценки математического ожидания и несмещенной дисперсии не применимы в случае цензурированных данных, необходимо получить эти оценки. Запишем определение математического ожидания $\mu=M\xi$ случайной величины ξ и соответствующей ей функции надежности $S(t)=1-F(t)$:

$$\mu=M\xi=\int_{-\infty}^{+\infty} t dF(t).$$

Таким образом, для вычисления математического ожидания в случае данных типа времени жизни необходима оценка $\hat{S}(t)$ для функции надежности $S(t)$. Для этого воспользуемся непараметрической оценкой функции надежности Абдушукурова $\hat{S}^A(t)$ [8], т.к. данная оценка, по сравнению с оценкой Каплана–Мейера [9], лишена недостатка и при любой степени цензурирования $\exists t:-\infty<t<\infty:\hat{S}^A(t)=0$.

Для выборки $X_i, i=1,2$ оценка Абдушукурова $\hat{S}_i^A(t)$ функции надежности $S_i(t)$ выглядит следующим образом:

$$\hat{S}_i^A(t)=\begin{cases}1, t<Z_{(i1)},\\ \left(\dfrac{n_i-j}{n_i}\right)^{R(t)}, Z_{(ij)}\le t<Z_{(ij+1)}, 1\le j\le n_i-1,\\ 0, t\ge Z_{(in_i)}.\end{cases}$$

где $R(t)=-\ln\left(\hat{S}_i^{Br}(t)\right)\Big/\sum_{j=1}^{n_i}\frac{I\left(Z_{(ij)}\le t\right)}{n_i-j+1}$, $\hat{S}_i^{Br}(t)=\exp\left(-\sum_{j=1}^{n_i}\frac{\delta_{(ij)}I\left(Z_{(ij)}\le t\right)}{n_i-j+1}\right), -\infty<t<+\infty.$

В соотношении выше используемая величина $\hat{S}_i^{Br}(t)$ является непараметрической оценкой функции надежности Бреслоу [10] для выборки $X_i, i=1,2$.

В итоге оценка математического ожидания в случае данных типа времени жизни может быть вычислена следующим образом:

$$\hat{\mu}_i^C=\sum_{j=1}^{n_i}Z_{(ij)}\left(\hat{S}_i^A(t-\varepsilon)-\hat{S}_i^A(t+\varepsilon)\right),\ \varepsilon\to 0.$$

Воспользуемся свойством [11] дисперсии случайной величины $s^2=D\xi$ для случайной величины ξ:

$$s^2=D\xi=M\xi^2-\left(M\xi\right)^2=\int_{-\infty}^{+\infty}t^2dS(t)-\left(\int_{-\infty}^{+\infty}tdS(t)\right)^2.$$

Тогда оценка несмещенной дисперсии случайной величины в случае данных типа времени жизни может быть вычислена следующим образом:

$$\left(s_i^C\right)^2=\frac{n_i}{n_i-1}\left(\sum_{j=1}^{n_i}Z_{(ij)}^2\left(\hat{S}_i^A(t-\varepsilon)-\hat{S}_i^A(t+\varepsilon)\right)-\left(\hat{\mu}_i^C\right)^2\right), \varepsilon\to 0.$$

Таким образом, если мы вместо величин $\hat{\mu}_i, s_i^2$ в приведенных соотношениях для критериев Стьюдента и Крамера–Уэлча будем использовать величины $\hat{\mu}_i^C, \left(s_i^C\right)^2$, то получим модифицированные статистики критериев Стьюдента $\hat{S}_t$ и Крамера–Уэлча $\hat{S}_{KW}$ для данных типа времени жизни. В дальнейшем речь будет идти о модифицированных критериях.

2.5. Распределение статистик модифицированного критерия Крамера–Уэлча

Для проведения процедуры проверки статистической гипотезы необходимо сравнить значение статистики критерия S_n с некоторым критическим значением S_α при фиксированной ошибке первого рода α. На основании этого сравнения делается вывод об отклонении или принятии статистической гипотезы. Критическое значение S_α берется из распределения статистики критерия при справедливости нулевой гипотезы $G(S_n\,|\,H_0)$ (в данной работе нулевая гипотеза – это равенство математических ожиданий), которое при $n\to\infty$ стремится к предельному распределению $G(S\,|\,H_0)$. На практике очень часто бывает, что при рассмотрении выборки объемом n распределение $G(S_n\,|\,H_0)$ еще существенно отличается от предельного распределения $G(S\,|\,H_0)$, поэтому использование предельного распределения $G(S\,|\,H_0)$ чревато неправильно сделанным выводом. Для определения степени близости распределения $G(S_n\,|\,H_0)$ к распределению $G(S\,|\,H_0)$ производят исследование скорости сходимости распределения статистик [12]. Покажем на рис. 1 и 2 для модификаций критериев Стьюдента и Крамера–Уэлча, что с ростом объемов выборок n_1, n_2 распределение статистик $G(S_n\,|\,H_0)$ приближается к $G(S\,|\,H_0)$.

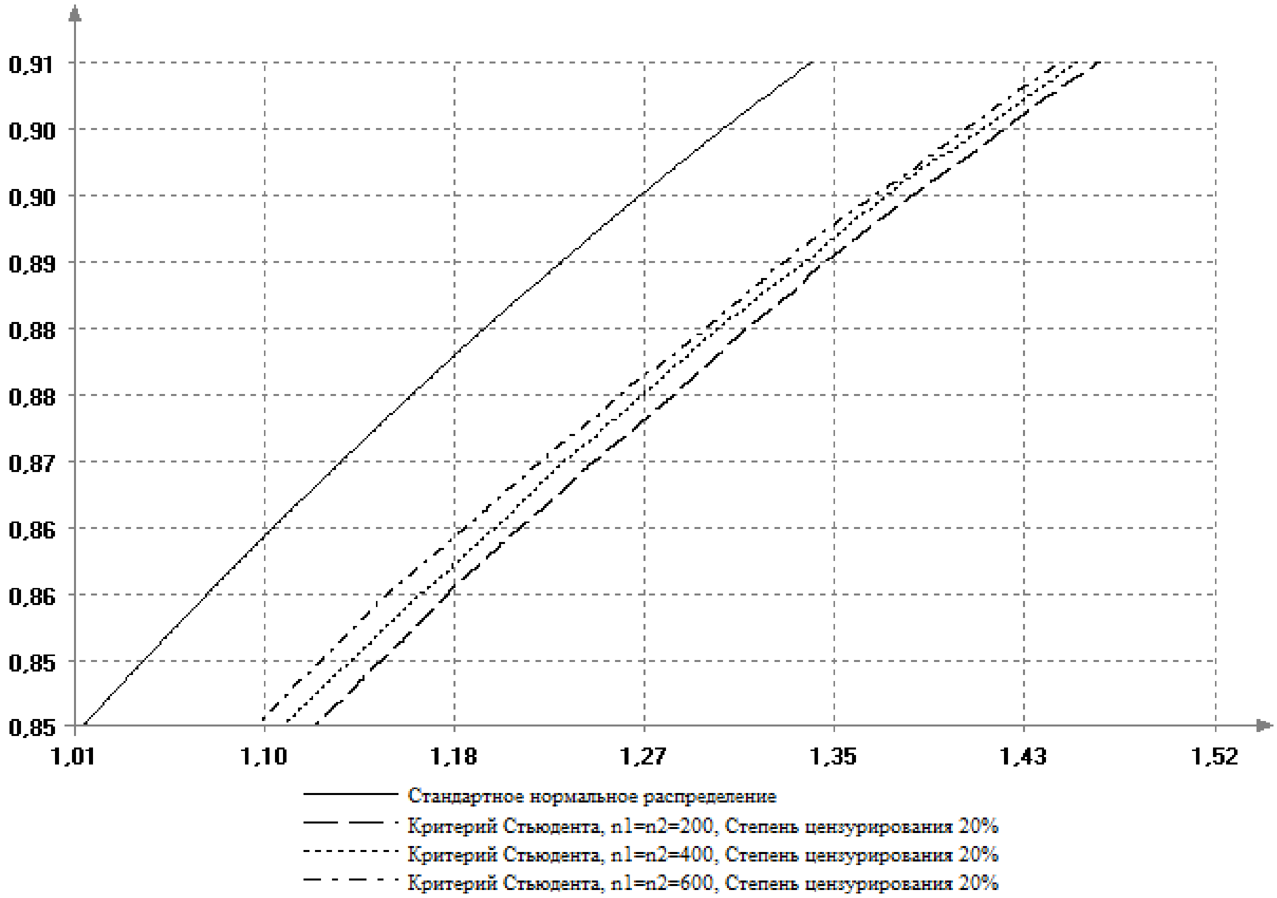

Рис. 1. Распределения статистик модифицированного критерия Стьюдента $G(\hat{S}_t \mid H_0)$ для разных объемов выборок n_1, n_2

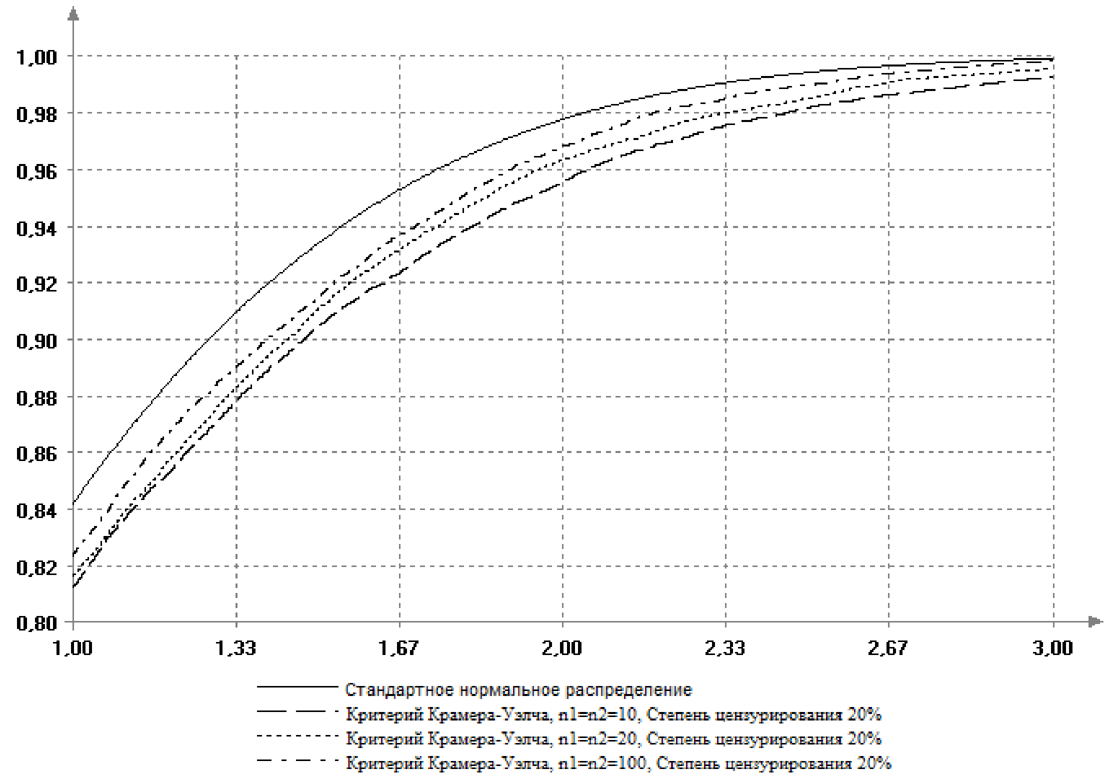

Рис. 2. Распределения статистик модифицированного критерия Крамера–Уэлча $G(\hat{S}_{KW} \mid H_0)$ для разных объемов выборок n_1, n_2

Как видно из рисунков выше, с ростом объемов выборок n_1, n_2 распределения $G(\hat{S}_t \mid H_0)$ и $G(\hat{S}_{KW} \mid H_0)$ становятся всё ближе к соответствующим предельным распределениям. Отметим тот факт, что в случае модифицированного критерия Стьюдента с ростом объемов выборок предельное распределение асимптотически стремится к стандартному нормальному распределению, именно поэтому были рассмотрены большие объемы выборок n_1, n_2.

Таким образом, в данной главе были рассмотрены классические критерии однородности математических ожиданий Стьюдента и Крамера–Уэлча, представлены оценки математического ожидания и несмещенной дисперсии в случае данных типа времени жизни с использованием оценки функции надежности Абдушукурова, а также показано, что с ростом объемов выборок распределения статистик критериев приближаются к соответствующим предельным распределениям.

3. Исследование мощности модифицированных критериев

Мощностью критерия называется вероятность правильно отвергнутой нулевой гипотезы, когда нулевая гипотеза не верна [13]. В подавляющем большинстве случаев на практике для критерия со статистикой S_n неизвестно распределение статистик при справедливости альтернативной гипотезы $G(S_n \mid H_1)$. В этом случае поведение функции мощности можно исследовать с помощью рассмотрения различных классов близких альтернативных гипотез. На основании полученных результатов делаются рекомендации о предпочтительности использования критерия на практике.

Покажем поведение мощности модифицированных критериев однородности математических ожиданий Стьюдента и Крамера–Уэлча при конкурирующих гипотезах с различной степенью близости, а также при конкурирующих гипотезах, для которых выполняются или нарушаются априорные предположения (для критерия Стьюдента) о том, что данные должны быть распределены по нормальному закону. Конкурирующие гипотезы показаны в табл. 1. Для указанных конкурирующих гипотез будем использовать цензурирование случайного типа с распределением Вейбулла–Гнеденко $We(t;\lambda,2)$, т.е. $F^C(t;\lambda) = We(t;\lambda,2)$, а параметр λ будет определяться из соотношения $\int_{-\infty}^{+\infty} \left(1 - F^C(t;\lambda)\right) f(t)dt = 1-\gamma$, где $f(t) = \frac{dF(t)}{dt}$, $\gamma \in [0,1]$ – заданная степень цензурирования.

Таблица 1. Альтернативные гипотезы

Обозначение	**Альтернатива**	**Значения параметров**	
		λ	δ
H_1	$\mathrm{N}(t;\lambda,1) - \mathrm{N}(t;\lambda+\delta,1)$	0	0.01
H_2	$\mathrm{N}(t;\lambda,1) - \mathrm{N}(t;\lambda+\delta,1)$	0	0.05
H_3	$\mathrm{Exp}(t;\lambda) - \mathrm{Exp}(t;\lambda-\delta)$	1	0.01
H_4	$\mathrm{Exp}(t;\lambda) - \mathrm{Exp}(t;\lambda-\delta)$	1	0.05

Соотношения функций плотности и используемых обозначений для указанных распределений в табл. 1 представлены в табл. 2.

Таблица 2. Используемые законы распределения

№	Обозначение	Закон распределения	Функция плотности
1	$\mathrm{N}\left(t;\lambda,\sigma^2\right)$	Нормальное распределение	$f_N(t;\lambda,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(t-\lambda\right)^2}{2\sigma^2}\right),t\in R$
2	$\mathrm{Exp}\left(t;\lambda\right)$	Экспоненциальное распределение	$f_{Exp}(t;\lambda)=\lambda\exp\left(-\lambda t\right),t\geq 0$
3	$We\left(t;\lambda,a\right)$	Распределение Вейбулла–Гнеденко	$f_{We}(t;\lambda,a)=\frac{a}{\lambda}\left(\frac{t}{\lambda}\right)^{a-1}\exp\left(-\left(\frac{t}{\lambda}\right)^a\right),t\geq 0$

С учетом введенных обозначений, рассматриваемых распределений, получим результаты компьютерного моделирования с использованием имитационного метода Монте-Карло. Объем моделирования N составил 20 000 повторений (в таком случае распределение статистик критерия $G(S_n \mid H_0)$ будет иметь погрешность моделирования менее 0.01 [14]), объемы выборок $n_1 = n_2 = 500$, зафиксированная ошибка первого рода $\alpha = 0.05$. Результаты представим в табл. 3.

Таблица 3. Результаты компьютерного моделирования

Критерий	H_i	0 %	10 %	20 %	30 %	40 %	50 %
Модифицир. критерий Стьюдента	H_1	0.052	0.052	0.046	0.053	0.056	0.062
	H_2	0.123	0.128	0.112	0.115	0.108	0.107
	H_3	0.052	0.057	0.054	0.056	0.050	0.048
	H_4	0.133	0.122	0.108	0.105	0.090	0.077
Модифицир. критерий Крамера–Уэлча	H_1	0.053	0.055	0.053	0.058	0.054	0.062
	H_2	0.113	0.122	0.116	0.111	0.118	0.106
	H_3	0.050	0.058	0.052	0.049	0.059	0.051
	H_4	0.129	0.116	0.109	0.099	0.095	0.077

Из полученных результатов в табл. 3 можно сделать выводы, что оба критерия очень плохо различают основную и альтернативную гипотезы, когда разница между математическими ожиданиями составляет не более 0.01, даже при объеме выборок $n_1 = n_2 = 500$, т.к. мощности критериев на альтернативных гипотезах H_1 и H_3 близки к значению зафиксированной ошибки первого рода $\alpha = 0.05$. Зато рассматриваемые критерии лучше определяют альтернативу, если разница между математическими ожиданиями не превышает 0.05, это можно увидеть по результатам мощности на альтернативных гипотезах H_2 и H_4. Отметим тот факт, что с ростом степени цензурирования мощность критериев уменьшается, как это показано с помощью результатов на альтернативных гипотезах H_2 и H_4. Кроме того, тот факт, что мощности модифицированных критериев Стьюдента и Крамера–Уэлча для рассмотренных альтернативных критериев имеют различие около 1 %, позволяет сделать вывод, что на практике действительно стоит использовать без существенной потери в мощности непараметрический критерий Крамера–Уэлча вместо критерия Стьюдента, для которого необходимо провести сложную проверку нормальности данных.

4. Заключение

Для проверки однородности выборок существуют различные подходы, как, например, проверка гипотезы о равенстве математических ожиданий. Для проверки такой статистической гипотезы на практике часто используют параметрический критерий Стьюдента. Однако существует непараметрический аналог данного критерия – критерий Крамера–Уэлча, статистика которого распределена по стандартному нормальному закону. Для рассмотренных конкурирующих гипотез мощности критериев близки, что подтверждает факт предпочтительности критерия Крамера–Уэлча вместо критерия Стьюдента, для которого необходимо проверять гипотезу о нормальности данных. С ростом степени цензурирования мощность критериев убывает.

В работе приведены оценки математического ожидания и дисперсии для данных типа времени жизни на основе оценки функции надежности Абдушукурова, а также аналоги критериев Стьюдента и Крамера–Уэлча для данных типа времени жизни. Показано, что с ростом объемов выборок распределение статистик критериев приближается к соответствующим предельным законам. Также показано, что критерии плохо различают альтернативную гипотезу, когда расстояние между математическими ожиданиями не более 0.01 даже при объемах выборок $n_1 = n_2 = 500$.

Литература

1. *Филоненко П. А., Постовалов С. Н.* Исследование влияния закона распределения моментов цензурирования и степени цензурирования на мощность критериев однородности // Сибирский журнал индустриальной математики. 2014. Т. 17, № 3. С. 122–134.
2. *Student*. The probable error of a mean // Biometrika. 1908. № 6 (1). P. 1–25.
3. *Орлов А. И.* Эконометрика. Изд. 4-е, доп. и перераб. Ростов-на-Дону: Феникс, 2009. 572 с.
4. *Орлов А. И.* Прикладная статистика. М.: Экзамен, 2006. 671 с.
5. *Орлов А. И.*, *Жихарев В. Н., Цупин В. А.* Анализ динамики цен на продовольственные товары в Москве и Московской области// В сб.: Научные труды Рижского института мировой экономики. 1998. В. 2. С.19–25.
6. *Philonenko P., Postovalov S.* A power comparison of homogeneity tests for randomly censored data // Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control (AMSA 2013), Novosibirsk, 25–27 Sept. 2013. P. 227–237.
7. *Welch B. L.* The generalization of “Student`s” problem when several different population variances are involved // Biometrika. 1947. V. 34. P. 29–35.
8. *Abdushukurov A. A.* Communications in Statistics: Theory and Methods. 1998. V. 27, № 8. P. 1991–2012.
9. *Kaplan E. L., Meier P.* Nonparametric estimator from incomplete observation // J. Amer. Statist. Assoc. 1958. V. 53. P. 457–481.
10. *Breslow N.* J. Roy. Stat. Soc. Ser. A. 1972. V. 34. P. 216–217.
11. *Бекарева Н. Д.* Теория вероятностей: учеб. пособие. Изд-во НГТУ, 2007. 196 с.
12. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: [монография]: монография / *Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова*. Новосибирск: Изд-во НГТУ, 2011. 888 с.
13. *Лемешко Б. Ю., Лемешко С. Б.* Об устойчивости и мощности критериев проверки однородности средних // Измерительная техника. М: ФГУП «Рос. науч.-техн. центр информации по стандартизации, метрологии и оценке соответствия», 2008. № 9. С. 23–28.

14. *Лемешко Б. Ю., Постовалов С. Н.* Компьютерные технологии анализа данных и исследования статистических закономерностей: учеб. пособие. Новосибирск: НГТУ, 2004. 119 с.

Статья поступила в редакцию 22.02.2016

Филоненко Петр Александрович
аспирант кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, пр. Карла Маркса, 20), тел. (383) 346-06-00, e-mail: petr-filonenko@mail.ru.

Постовалов Сергей Николаевич
д.т.н., доцент кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, пр. Карла Маркса, 20), тел. (383) 346-06-00, e-mail: postovalov@ngs.ru.

The Student and Cramer–Welch tests for two-sample problem testing with lifetime data

P. Philonenko, S. Postovalov

The parametric Student test and the nonparametric Cramer–Welch test for two-sample problem of means equality are considered. These tests can be applied for two-sample problem testing. We propose statistics of the tests for the lifetime data case. We present that the distributions of the test statistic look forward to the limit distributions while sample sizes increase. Moreover, the results of the test power simulation are presented in this paper using the Monte-Carlo method. The alternative hypotheses of the test power simulation contain alternative hypotheses that breach of a priory conditions of the data normality.

Keywords: two-sample problem, Student test, Cramer–Welch test, lifetime data, parametric test, nonparametric, test power, test statistic distribution, Monte-Carlo method.