

Фурина Ксения Олеговна, студент

Факультет прикладной математики и механики,

Федеральное государственное бюджетное образовательное учреждение

высшего образования «Пермский национальный

исследовательский политехнический университет»

e-mail: darkcityfightclub@mail.ru

ПОСТРОЕНИЕ ТОЛЕРАНТНЫХ ИНТЕРВАЛОВ ДЛЯ РЕГРЕССИИ ПО НЕОДНОРОДНОЙ ВЫБОРКЕ

Аннотация: В статье рассматривается неоднородная выборка, а затем она разбивается на две однородные выборки и для каждой находятся толерантные интервалы.

Ключевые слова: толерантные интервалы, множественная регрессия, неоднородная выборка.

Abstract: The article discusses nonuniform sampling, and then it is split into two homogenous sample and are tolerant of each intervals.

Keywords: tolerance intervals, multiple regression, non-uniform sampling.

При медицинском количественном типе исследования результаты выдаются в форме цифр. При этом существует диапазон нормы, а также средние показатели. Референтное значение в анализах – это термин медицины, применяемый для оценки результатов при лабораторных исследованиях. Определяется оно как среднее значение определенного показателя. Эти данные получены путем обследования здоровой части населения.

Первым этапом является выборка населения. Результаты сводят к средним цифрам, вычисляя диапазон, в которых находятся референтные

значения. Допускается отклонение от нормальных показателей (в ту или другую сторону) на две стандартных единицы.

Референтное значение в анализах – это статистические данные, но не биологический закон. Как правило, диагностические выводы делаются не в соответствии с единичными показателями, а при оценке динамики изменений [3].

В математической статистике, учитывая определенные уровни объясняющих переменных, обычно применяют два статистических интервала: интервалы для средней объясняемой переменной (доверительные интервалы) и интервалы для одного или более будущего наблюдения (прогнозные интервалы). Толерантные интервалы являются другим типом статистического интервала.

Толерантный интервал – интервал, определяемый по выборке, относительно которого можно утверждать с уровнем доверия $(1 - \alpha)$, что он содержит, по крайней мере, указанную долю p совокупности. Границы статистического толерантного интервала называются статистическими толерантными границами. Уровень доверия $(1 - \alpha)$ - это вероятность того, что толерантный, определенный описанным методом, будет содержать не менее чем долю p совокупности. Наоборот, вероятность того, что толерантный интервал будет содержать менее чем долю p совокупности, есть α . Настоящий стандарт описывает методы определения односторонних (с верхней или нижней границей) и двусторонних (с верхней и нижней границами) статистических толерантных интервалов.

Толерантный интервал является функцией наблюдений выборки, то есть статистики. Приведенные в настоящем стандарте методы предполагают, что наблюдения в выборке независимы [1].

Рассмотрим выборку, которая неоднородна по признаку Группа.

Таблица 1 - Выборка для референтного значения определенного показателя

№	X	Y	Группа
1	-2,1549	4,622711	2
2	-2,1549	4,752102	2
3	-2,22185	5	2
4	-2,09691	4,377215	2
5	-2,09691	3,908592	2
6	-2,22185	5,000226	2
7	-2,11919	4,976671	1
8	-2,04576	4,488142	1
9	-2,22185	5,155275	1
10	-2,09691	5,075463	1
11	-2,1549	5	1
12	-2,04576	4,056447	1
13	-2,09691	4,141763	1
14	-2,1549	4,825082	2
15	-2,04576	4,031126	2
16	-2,20761	4,726776	2
17	-2,1549	4,647706	2
18	-2,09691	4,527875	2
19	-2,04576	3,965343	2
20	-2,23657	5,136622	2
21	-2	3,716254	2
22	-2,09691	3,871398	2
23	-2,1549	4,876605	2
24	-2,1549	4,48243	2
25	-2,04576	3,678882	2
26	-2,18709	4,791726	2
27	-2,12494	4,576053	2
28	-2,09691	3,968156	2
29	-2,22185	5,016432	2
30	-2,07058	4,495406	1
31	-2,04576	4,524941	1
32	-2,09691	4,680934	1
33	-2,22185	5,889076	1
34	-2	3,736954	1
35	-2,09691	4,967506	1
36	-2,04576	3,840482	1
37	-2,07058	4,577779	1
38	-2,04576	4,066624	1

Разобьём эту выборку на две подвыборки: по Группе 1 и Группе 2.

Таблица 2 – Сгруппированная выборка для референтного значения определенного показателя

№	X	Y	Группа
1	-2,1549	4,622711	2
2	-2,1549	4,752102	
3	-2,22185	5	
4	-2,09691	4,377215	
5	-2,09691	3,908592	
6	-2,22185	5,000226	
7	-2,1549	4,825082	
8	-2,04576	4,031126	
9	-2,20761	4,726776	
10	-2,1549	4,647706	
11	-2,09691	4,527875	
12	-2,04576	3,965343	
13	-2,23657	5,136622	
14	-2	3,716254	
15	-2,09691	3,871398	
16	-2,1549	4,876605	
17	-2,1549	4,48243	
18	-2,04576	3,678882	
19	-2,18709	4,791726	
20	-2,12494	4,576053	
21	-2,09691	3,968156	
22	-2,22185	5,016432	
23	-2,07058	4,495406	1
24	-2,04576	4,524941	
25	-2,09691	4,680934	
26	-2,22185	5,889076	
27	-2	3,736954	
28	-2,09691	4,967506	
29	-2,04576	3,840482	
30	-2,07058	4,577779	
31	-2,04576	4,066624	
32	-2,11919	4,976671	
33	-2,04576	4,488142	
34	-2,22185	5,155275	
35	-2,09691	5,075463	
36	-2,1549	5	
37	-2,04576	4,056447	
38	-2,09691	4,141763	

Оценим параметры квадратичной зависимости $Y = \alpha + \beta X + \gamma X^2 + \varepsilon$, где α, β, γ – параметры, ε – случайные отклонения (остатки модели).

Затем найдем нижнюю границу одностороннего толерантного интервала регрессии сначала для Группы 2, а затем для Группы 1.

Полагая $X_1 = X^2$, получим множественную регрессию $Y = \alpha + \beta X + \gamma X_1 + \varepsilon$ [2].

1. Будем полагать доля $p = 0,9$, уровень доверия $\gamma = 1 - \alpha = 0,95$, объем 1 выборки $n = 22$.

2. Находим оценку вектора параметров регрессии

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = (X^T X)^{-1} \cdot X^T Y = \begin{pmatrix} 46913,4 & 44109,7 & 10358,7 \\ 44109,7 & 41484,7 & 9744,78 \\ 10358,7 & 9744,78 & 2289,65 \end{pmatrix} \cdot \begin{pmatrix} 98,499 \\ -210,9 \\ 451,966 \end{pmatrix} = \begin{pmatrix} -15,78 \\ -12,639 \\ -1,4747 \end{pmatrix}, \text{ т.е.}$$

$$\hat{Y} = -15,78 - 12,639X - 1,4747X_1, \quad R^2 = 0.837.$$

3. Находим оценку дисперсии случайных отклонений $s^2 = \frac{(Y - X\hat{\beta})^T (Y - X\hat{\beta})}{22 - 3} = 0,0385$

4. Для заданной доли $p = 0,9$ находим квантиль стандартного нормального распределения $z_{0,9} = 1,2816$.

5. Для выбранного вектора-столбца (наблюдение 1, Группа 2) $x = \begin{pmatrix} 1 \\ -2,1549 \\ 4,6436 \end{pmatrix}$

находим

$$d = (x^T (X^T X)^{-1} x)^{1/2} = \left[(1 \quad -2,1549 \quad 4,6436) \cdot \begin{pmatrix} 46913,4 & 44109,7 & 10358,7 \\ 44109,7 & 41484,7 & 9744,78 \\ 10358,7 & 9744,78 & 2289,65 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2,1549 \\ 4,6436 \end{pmatrix} \right]^{1/2} = 0,2801$$

$$t_{n-m; 1-\alpha}(z_p/d) = t_{19; 0,95}\left(\frac{1,2816}{0,2801}\right) = t_{19; 0,95}(4,58) = 7,146,$$

$$k(d) = d \times t_{19; 0,95}(4,58) = 0,2801 \cdot 7,146 = 2,$$

$$x^T \hat{\beta} = (1 \quad -2,1549 \quad 4,6436) \cdot \begin{pmatrix} -15,78 \\ -12,639 \\ -1,475 \end{pmatrix} = 4,608.$$

6. Для выбранного вектора-столбца $x = \begin{pmatrix} 1 \\ -2,1549 \\ 4,6436 \end{pmatrix}$ находим нижнюю границу

одностороннего толерантного интервала

$$L(x) = x^T \hat{\beta} - k(d(x))S = 4,608 - 2 \cdot 0,196 = 4.215.$$

Повторим пункты 1- 6 для наблюдения 23 выборки.

1. Будем полагать доля $p = 0,9$, уровень доверия $\gamma = 1 - \alpha = 0,95$, объем 1 выборки $n = 16$.

2. Находим оценку вектора параметров регрессии

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = (X^T X)^{-1} \cdot X^T Y = \begin{pmatrix} 73613,5 & 69463,5 & 16370,4 \\ 69463,5 & 65563,6 & 15455,2 \\ 16370,4 & 15455,2 & 3644,19 \end{pmatrix} \cdot \begin{pmatrix} 73,67 \\ -154,59 \\ 324,68 \end{pmatrix} = \begin{pmatrix} -90,62 \\ -82,364 \\ -17,598 \end{pmatrix}, \text{ т.е.}$$

$$\hat{Y} = -90,62 - 82,364X - 17,598X1, \quad R^2 = 0.837.$$

3. Находим оценку дисперсии случайных отклонений $S^2 = \frac{(Y - X\hat{\beta})^T (Y - X\hat{\beta})}{22 - 3} = 0,0919$

4. Для заданной доли $p = 0,9$ находим квантиль стандартного нормального распределения $z_{0,9} = 1,2816$.

5. Для выбранного вектора-столбца (наблюдение 1, Группа 1) $x = \begin{pmatrix} 1 \\ -2,0706 \\ 4,28731 \end{pmatrix}$

находим

$$d = (x^T (X^T X)^{-1} x)^{1/2} = \left[(1 \quad -2,0706 \quad 4,28731) \cdot \begin{pmatrix} 73613,5 & 69463,5 & 16370,4 \\ 69463,5 & 65563,6 & 15455,2 \\ 16370,4 & 15455,2 & 3644,19 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2,0706 \\ 4,28731 \end{pmatrix} \right]^{1/2} = 0,2909$$

$$t_{n-m; 1-\alpha}(z_p/d) = t_{13; 0,95} \left(\frac{1,2816}{0,2909} \right) = t_{13; 0,95}(4,41) = 7,357,$$

$$k(d) = d \times t_{19; 0,95}(4,58) = 0,2909 \cdot 7,357 = 2,14,$$

$$x^T \hat{\beta} = (1 \quad -2,0706 \quad 4,28731) \cdot \begin{pmatrix} -90,62 \\ -82,364 \\ -17,598 \end{pmatrix} = 4.472.$$

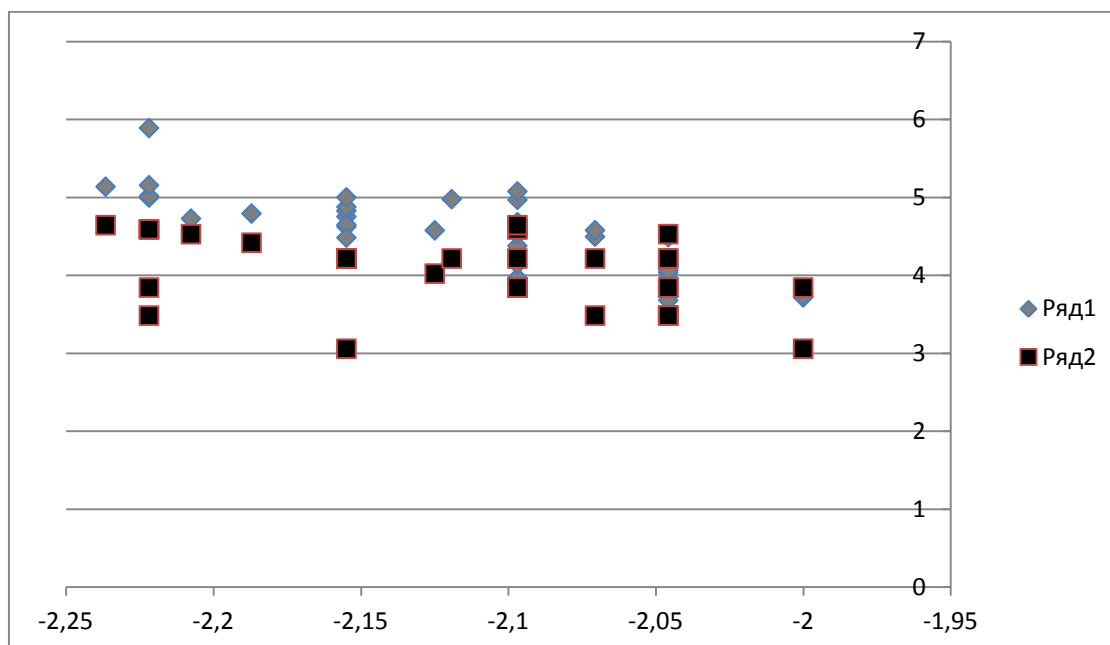
6. Для выбранного вектора-столбца $x = \begin{pmatrix} 1 \\ -2,0706 \\ 4,28731 \end{pmatrix}$ находим нижнюю границу одностороннего толерантного интервала

$$L(x) = x^T \hat{\beta} - k(d(x))S = 4,472 - 2,14 \cdot 0,303 = 4.215.$$

Таблица 3 - Результаты вычислений

№	X	Y	L
1	-2,1549	4,622711	4,21544
2	-2,1549	4,752102	4,21544
3	-2,22185	5	4,589782
4	-2,09691	4,377215	3,842352
5	-2,09691	3,908592	3,842352
6	-2,22185	5,000226	4,589782
7	-2,1549	4,825082	4,21544
8	-2,04576	4,031126	3,482807
9	-2,20761	4,726776	4,527215
10	-2,1549	4,647706	4,21544
11	-2,09691	4,527875	3,842352
12	-2,04576	3,965343	3,482807
13	-2,23657	5,136622	4,642461
14	-2	3,716254	3,05749
15	-2,09691	3,871398	3,842352
16	-2,1549	4,876605	4,21544
17	-2,1549	4,48243	4,21544
18	-2,04576	3,678882	3,482807
19	-2,18709	4,791726	4,417336
20	-2,12494	4,576053	4,021211
21	-2,09691	3,968156	3,842352
22	-2,22185	5,016432	4,589782
23	-2,07058	4,495406	4,21544
24	-2,04576	4,524941	4,21544
25	-2,09691	4,680934	4,589782
26	-2,22185	5,889076	3,842352
27	-2	3,736954	3,842352
28	-2,09691	4,967506	4,589782
29	-2,04576	3,840482	4,21544
30	-2,07058	4,577779	3,482807
31	-2,04576	4,066624	4,527215
32	-2,11919	4,976671	4,21544
33	-2,04576	4,488142	3,842352
34	-2,22185	5,155275	3,482807
35	-2,09691	5,075463	4,642461
36	-2,1549	5	3,05749
37	-2,04576	4,056447	3,842352
38	-2,09691	4,141763	4,21544

Рисунок 1 – Нижняя граница для референтного значения



Библиографический список:

1. ГОСТ Р ИСО 16269-6-2005. Статистические методы. Статистическое представление данных. Определение статистических толерантных интервалов. (УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 30 июня 2005 г. N 171-ст) М.: Стандартинформ, 2005.
2. Ш. Закс, Теория статистических выводов, М.: МИР, 1975, 776 с.
3. FB.ru Интернет журнал [Электронный ресурс] Референтные значения – что это такое? <http://fb.ru/article/136705/referentnyie-znacheniya---chto-eto-takoe-chto-oznachaet-referentnoe-znachenie> (дата обращения: 10.12.2016).