УДК:658.382-012.

МЕТОДИКА ОБРАБОТКИ РЕЗУЛЬТАТОВ НАУЧНО – ИССЛЕДОВАТЕЛЬСКИХ РАБОТ

Юлдашев Орунбай Рахмонбердиевич
кандидат технических наук,
доцент Института гражданской защиты
при Академии МЧС РУз
Рахимжанова Дильфуза Хасановна
преподаватель Института гражданской защиты
при Академии МЧС РУз
Мансуржанова Шохсанам Расулжонкизи
магистрант второго курса
Института гражданской защиты
при Академии МЧС РУз

Аннотация: *В данной статье рассматривается вопросы регулирования отношений, связанных при исследовании количественных выражений того или иного признака, а также закономерность распределения частот результатов измерений распределения.*

Ключевые слова: *стандартное, сигма, сравнение средних, средних арифметических, значение величины, доверительной вероятностью прямоугольного треугольника.*

Annotation: *This article discusses the issues of regulation of relations related to the study of quantitative expressions of a particular trait, as well as the regularity of the frequency distribution of the results of distribution measurements.*

Key words: *standard, sigma, comparison of means, arithmetic means, value of quantity, confidence probability of a right triangle.*

Введение. При исследовании количественных выражений того или иного признака часто обнаруживается, что результаты группируются вокруг некоторого центрального значения. Результаты, близкие к центральному значению, встречаются чаще. Результаты же, отличные от центрального значения встречаются реже и тем реже, чем больше это отличие. Эта закономерность распределения частот результатов измерений может быть определена количественно, так называемым, законом нормального распределения[3,2].

Для практического использования важны следующие свойства нормального распределения:

1. Центральное значение точнее всего определяется средней арифметической, обозначаемой буквой *M*или латинскойбуквой с черточкой наверху (x, y и т. д.). Чтобы найти *M*, как известно, делят сумму полученных результатов на их число:

$$M = (2+4+2+3) : 5 = 2{,}2$$

В общем виде формула вычисления средней выглядит так:

$$\bar{x} = \frac{X1+X2+\cdots.....+\Pi n}{n} = \frac{i=1\ Xi}{n} = \frac{\Sigma x}{n}.(1)$$

Здесь посредством x_1, x_2..., x_n обозначены результаты измерений. СимволΣx_iили, проще Σ_x обозначает суммирование всех *n* результатов [2].

2. Мерой разброса результатов около средней является среднее квадратичное (стандартное) отклонение, обозначаемая греческой буквой σ (сигма):

$$\sigma = \sqrt{\frac{(X1-X)2+(X2-X)2+\cdots...+(Xn-X)2}{n-1}} = \sqrt{\frac{\Sigma(X-X)2}{n-1}},(2)$$

σ можно определить и по другой, преобразованной формуле:

$$\sigma = \sqrt{\frac{1}{n-1}[\Sigma x^2 - \frac{(Ex)^2}{n}]}(3)$$

3. Далее необходимо отметить, что получить истинное значение средней нельзя и полученный результат *M* сопряжен с ошибкой. Ошибка

среднего зависит от ряда факторов, в том числе и от количества наблюдений. Очевидно, чем больше наблюдений, тем менее ошибка, и наоборот. Ошибка средней *m* вычисляется по формуле:

$$\mathbf{m} = \frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}} \quad (4)$$

Из формулы (4) видно, что ошибка среднего в корень квадратный из *n* раз меньше среднего стандартного отклонения. К. Гаусс установил, что ошибки средних подчиняются нормальному закону. Причем этому закону подчиняются ошибки средних и тех признаков, распределения результатов, которых значительно отличаются от нормального [4,8].

Из этого следует, что $M \pm tm$ представляет собой интервал, в котором с доверительной вероятностью (вероятностью, с которой мы верим результату) ($100—P_{(t)}$) % (или в долях единицы $1—P^{(t)}$) заключено истинное значение средней. Вероятность $P_{(t)}$ носит название уровня значимости и означает вероятность того, что истинное значение средней будет лежать за пределами интервала $M \pm tm$.

Необходимо отметить, что в интервал $M \pm 1{,}96m$ с доверительной вероятностью 0,95 включено истинное значение средней. Это заключение ошибочно в 5% случаев, т. е. P=0,05. Для большинства исследований доверительная вероятность 0,95 считается достаточной.

Методы и материалы.Важной особенностью технических и социально-гигиенических исследований является часто ограниченное число случаев (наблюдений, опытов). Здесь важно помнить, что при числе наблюдений меньше 30, для получения одной и той же доверительной вероятности, надо брать число *t* большим, и тем большим, чем меньше *n*. В этом случае мы используем данные, так называемой таблицы (табл. 1). Величина *t* зависит от доверительной вероятности (или от уровня значимости *P*) и от, так называемого числа степеней свободы *К* в данном случае равным $n—1$. Так при n=10 и при P=0,05 t= 2,26.[8].

Как известно, научное исследование предполагает наличие не менее 2 групп или одной группы, но рассмотренной по нескольким стадиям. Задача исследования состоит в установлении отличия между собой данных этих групп. Последнее дает возможность утверждать о наличии эффекта, например, от нового средства или метода профилактики. Наиболее простым способом сравнение данных является сравнение средних арифметических.

Основная идея статистической методики сравнения средних состоит в том, что это сравнение производится с учетом их ошибок. Сравнение производится между средними двух групп. Если групп несколько, то процесс сравнения производится поочередно.

Сравнение средних, когда числа наблюдений в группах равны (n_1=n_2). В этом случае вычисление производится по формуле:

$$t = \frac{M_1 M_2}{m\text{общ}},(5)$$

где $m \text{ общ} = \sqrt{{m_1}^2 + {m_2}^2}$ (6)

или, в общем виде:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} = \frac{M_1 - M_2}{\sqrt{{m_1}^2 + {m_2}^2}} \quad .(5^1)$$

Иными словами, здесь t является отношением разности средних и их суммарной ошибки. Практически можно считать, что если это отношение больше 2 (t>2), при суммарном числе наблюдений больше 30 ($n_1 + n_2$>30), то разница между средними достоверна. В этом случае P<0,05 (табл. 1). Здесь число степеней свободы к = $n_1 + n^2$—2.[10]

Таблица 1

Значение величины t при малом числе наблюдений

k	$P_{(t)}$				
	0,10	0,05	0,02	0,01	0,001

1	6,31	12,71	31,82	63,66	636,62
2	2,92	4,30	6,97	9,93	31,60
3	2,35	3,18	4,54	5,84	12,94
4	2,13	2,78	3,75	4,60	8,61
5	2,02	2,57	3,47	4,03	6,86
6	1,94	2.45	3,14	3,70	5,96
7	1,90	2,37	3,00	3,50	5,40
8	1,86	2,30	2,90	3,36	5,04.
9	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,49
12	1 ,78	2,18	2.68	3,06	4,32
13	1,77	2,16	2,65	3,01	4,22
14	1,76	2,14	2,62	2,98	4, 14
15	1,75	2,13	2,60	2,95	4,07
16	1,75	2,12	2,58	2,92	4,02
17	1,74	2,11	2,57	2,90	3,97
18	1,73	2,10	2,55	2,88	3,92
19	1,73	2,09	2,54	2,86	3,38
20	1,72	2,09	2,53	2,85	3,85
21	1,72	2,08	2,52	2,83	3,82
22	1,72	2,07	2,51	2,82	3,79
23	1,71	2,07	2,50	2,81	3,77
24	1,71	2,06	2,49	2,80	3,75
25	1,71	2,06	2,49	2,79	3,72
26	1,71	2,06	2,48	2,78	3,71
27	1,70	2,05	2,47	2,77	3,69
28	1,70	2,05	2,47	2,76	3,67
29	1,70	2,05	2,46	2,76	3,66

30	1,70	2,04	2,46	2,75	3,05
40	1,68	2,02	2,42	2,70	3,55
60	1,67	2,00	2,39	2,66	3,46
120	1,66	1,98	2,36	2,62	3,37
360	1,65	1,96	2,33	2,58	3,26

Результаты. Установлено, что между размахом и стандартным отклонением существует зависимость:

$$\sigma = \frac{R}{A} = \frac{X_{max} - X_{min}}{A},$$ где (7)

А-табличный коэффициент в зависимости от числа наблюдений n (табл. 2).

Аналогично $m = \frac{R}{B}$, где (8)

В—табличный коэффициент (В=А $\sqrt{n}$)

Таким образом, самая сложная часть формулы — это вычисление ошибок, сводится к простой процедуре — отнять и разделить. Далее, вычисление суммарной ошибки $m_{общ}$, можно еще более упростить. Заметим, что квадрат суммарной ошибки равен сумме квадратов обеих ошибок:

$$m^2_{общ} = m_1^{\ 2} + m_2^{\ 2},$$

Последнее означает, что $m_{общ}$ можно представить как гипотенузу прямоугольного треугольника, катетами которого являются ошибки. Тогда вычислив ошибки по размаху и отложив их на миллиметровой бумаге, при помощи обычной линейки (лучше, если она изготовлена из прозрачного материала) определяем величину гипотенузы. Последняя будет равна$m_{общ}$. Действия возведения ошибок в квадрат, сложение их и извлечение корня квадратного из полученной суммы здесь исключаются. Точность результатов практически достаточна.

Сравнение средних, когда числа наблюдений в группах не равны(n_1 = $n_2^{/}$). В этом случае вычисления производятся по более сложной формуле:

$$\mathbf{t} = \frac{M_1 - M_2}{\sqrt{(n_1 - 1)\,\sigma_1^2 + (n_2 - 1)\,\sigma_2^2}} \cdot \sqrt{\frac{n_1 \cdot n_2\,(n_1 + n_2 - 2)}{n_1 + n_2}} \quad (9)$$

Мы предлагаем более простую и доступную формулу, вычисления по которой дают результаты, достаточно близкие к результатам, полученным по формуле (9):

$$t = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{n_2} + \frac{\sigma_2^2}{n_1}\sigma_2}} \qquad (10)$$

Формула (10) внешне напоминает формулу (5') с той разницей, что ошибки здесь получаются делением на «соседнее» число наблюдений (перекрестное деление).

Вычисления по этой формуле более доступны, чем по формуле (9).

Приведем 2 примера.

Пример 1. У студентов проводились исследования пульса до и после сдачи экзамена (по данным кафедры - зимняя экзаменационная сессия учебного года). Данные частоты пульса приведены ниже[5].

Частота пульса до экзамена….	80	82	90	94	95	105	107	110	117	120
После экзамена...	69	73	75	79	85	80	92	94	97	100

Частота пульса,в среднем до экзамена составила 100 ударов в 1 мин, после экзамена 85.

Можно ли на основании этих данных считать, что после экзаменов частота пульса снижается и приближается к норме?

Здесь число случаев в обеих группах равны. Точнее, здесь одна группа, рассмотренная по двум стадиям до и после экзаменов. Вычисления производим по формуле (5'), для чего необходимо определить ошибки. Максимальный результат до экзамена составляет 120, минимальный 80. Размах - разница между максимальным и минимальным значением - равен 40. Последнюю цифру делим па коэффициент В, равный 9,70 (строка 10), и получаем ошибку средней, равную 4,12. Аналогичные вычисления производим и с результатами, полученными после экзаменов. Здесь ошибка составила 3,20. По способу «прямоугольного треугольника» определим суммарную ошибку разности. Она равна 5,22. Определим величину *t*:

$$t = \frac{M^1 - M^2}{\sqrt{m_1{}^2 + m_2{}^2}} = \frac{100-85}{\sqrt{4{,}12^2+3{,}20^2}} = \frac{15}{5{,}22} = 2{,}87$$

От величины*t* по таблице 1 переходим к *P* в зависимости от числа степеней свободы $K=n_1+n_2$—2=10+10—2=18. Из табл. 1 видно, что при *P=0,02 t=2,55*, а при*P=0,01 t=2,88.*

Таблица 2.

Значение коэффициентов А и В (в зависимости от числа наблюдений (n), на которое надо делить размах (R), чтобы получить соответственно значения среднего квадратичного отклонения (σ) и ошибки средней (m).

п	А	В	п	А	В
1	_	—	120	5,15	56,2
2	1,13	1,60	140	5,26	62,3
3	1,69	2,93	160	5,35	67,6
4	2,06	4,12	180	5,43	73,0
5	2,33	5,20	200	5,50	77,8
6	2,53	6,20	220	5,57	82,6
7	2,70	7,16	240	5,61	87,0
8	2,85	8,05	260	5,68	91,7
9	2,97	8,90	280	5,72	95,7
10	3,08	9,70	300	5,77	100,0

11	3,17	10,5	320	5,80	103,8
12	3,26	1 I ,2	340	5,84	107,9
13	3,34	12,0	360	5,88	111,5
14	3,41	12,7	380	5,92	115,2
15	3,47	13,4	400	5,94	118,8
16	3,53	14,1	420	5,98	122,6
17	3,59	14,8	440	6,00	125,9
18	3,64	15,4	460	6,02	129,2
19	3,69	16,1	480	6,06	132,8
20	3,74	16,7	500	6,09	136,0
22	3,82	17,9	520	6,12	139,3
24	3,90	19,0	540	6,13	142,5
26	3,96	20,2	560	6,14	145.6
28	4,03	21,2	580	6,17	148,6
30	4,09	22,4	600	6,18	151,5
32	4,14	23,4	620	6,21	154,6
34	4,19	24,6	640	6,23	157,7
36	4,24	25,5	660	6,26	160,8
38	4,28	26,4	680	6,27	163,4
40	4,32	27,3	700	6,28	166,4
50	4,50	31,8	750	6,33	173,3
60	4,64	35,9	800	6,34	177,9
70	4,76	39,8	850	6,37	186,6
80	4,85	43,3	900	6,43	193,0
90	4,94	46,9	950	6,47	199,2
100	5,01	50,1	1000	6,48	204,9

Примечание. *В случаях, когда числа наблюдений в исследованиях отличаются от табличных, коэффициенты А и Вполучаются методом интерполяции.*

Значение 2,87 занимает промежуточное положение и можно утверждать, что P < 0,02. Последнее означает, что более 98 шансов из 100 против 2 за то, что средние отличаются друг от друга. Следовательно, можно считать разницу достоверной и обусловленной экзаменом как стрессовой ситуацией, сила которой снимается после сдачи экзамена. Наглядно это можно представить в виде табл. 3 [6].

Таблица 3.

Среднее значение частоты пульса у студентов

до и после сдачи экзаменов

Статистические показатели	**Частота пульса**	
	до экзамена	**после экзамена**
M ±m	100,0±4,12	85,0±3,20
n	10	10
P	*P*<0,02	

Пример 2. Имеются 2 группы *x* и *y*.

x	3,0	5,0	4,0	3,5		
y	3,0	4,0	7,0	7,0	10,0	11,0

Необходимо определить имеется ли достоверная разница между средними.

Здесь мы имеем случай, когда n_1 / n_2, следовательнорассчеты надо вести по формуле (10). Вначале вычисляем средние, затем стандартные отклонения и ошибки средних по размаху и определяем *t:*

$$t = \frac{x - y}{\sqrt{\frac{\sigma_X{}^2}{n_Y} + \frac{\sigma_Y{}^2}{n_X}}} = \frac{3{,}75 - 7{,}00}{\sqrt{\frac{\left(\frac{5-3}{2{,}06}\right)^2}{6} + \frac{\left(\frac{11-3}{2{,}53}\right)}{4}}} =$$

$$= \frac{3{,}25}{\sqrt{\frac{0{,}94}{6} + \frac{9{,}99}{4}}} = 1{,}99$$

От t по табл. 1 переходим к P. На пересечении $K = n_1 + n_2 — 2 = 4 + 6 — 2 = 8$ и $P = 0,10$ имеет $t = 1,86$, а $K=8$ и $P = 0,05$ $t= 2,30$. Иными словами, P лежит между 0,10 и 0,05, т. е. $P > 0,05$. Уровень значимости весьма высок и нельзя утверждать о наличии разницы между средними. Здесь рассуждать надо следующим образом: или же разницы нет в действительности или же она есть, но не может быть определена за счет больших ошибок, обусловленных малым числом наблюдений. Более общее заключение может быть изложено следующим образом: на основании этих данных нельзя утверждать о наличии между средними достоверной разницы. Решить вопрос об увеличении числа наблюдений должен сам исследователь.

Сравнительная характеристика данных, полученных по различным формулам и ускоренной методике в приложении к примеру 2 приведены в табл. 4.

Таблица 4.

Статистические показатели	**Рассчитано по**			
	формуле (9)		ускоренной методике	формуле (5,5')
	по форм. (2)	по размаху		
t	1,97	1,96	1,99	2,36
P	$P > 0,05$	$P > 0,05$	$P > 0,05$	$P < 0,05$

Выводы: Поразительна точность совпадения результатов, рассчитанная по ускоренной методике с вычислением σ итпо размаху с результатами, рассчитанными по формулам (9) и (2).

Что касается случаев вычисления по формуле (5 и 5'), когда n_1 / n_2, (что иногда, к сожалению, имеет место), то часто получаются значения P, отличные от истинных и исследователь может прийти к неправильным выводам, что показано в вышеприведенной таблице: по этой формуле получаем $t = 2,36$ и $P < 0,05$; на основании чего исследователь может сделать вывод о наличии достоверной разницы.

В заключение отметим, что изложенная методика статистической обработки данных является лишь средством, с помощью которого исследователь производит объективные оценки в случаях, когда при попытках объяснить наблюдаемые факты могут возникнуть внутренние противоречия. Одно из основных требований, предъявляемых к данной методике, состоит в том, что посредством ее можно выявлятьзакономерности, лежащие в основе признака пли явлении, характеризуемого данным экспериментальным статистическим материалом.

Список использованной литературы:

1. Исторический опыт, современные проблемы и перспективы образовательной и научной деятельности в области пожарной безопасности. Сборник тезисов, докладов, материалов международной научно-практической конференции. Москва 28-29 октября 2018г.435 стр.

2. Ускоренные испытания на надёжность технических систем Новосибирск.МТ-4,2014г., часть 2, 344 стр.

3. В.В.Кроличенко. Методика оценки риска последствий аварий на гидротехнических сооружениях напорного типа с применением аэрогеодезических технологий идентификации их устойчивости в экстремальных ситуациях. (Автореферат на соиск.учён.степ.канд.техн.наук Самара 2012г. – 12с.).

4. Е.Н.Беллендир, Д.А.Ивашинцов и др. Вероятностные методы оценки надёжности грунтовых гидротехнических сооружений. – СПб.изд-ва ОАО «ВНИИГ им.Б.Е.Веденеева» - 2003г. 1-253с.

5. Инженерный анализ последствий землетрясений в Японии и США М.Госстойиздать. 1961-193с.

6. Кудрин Ю.М. и др. Применение автоматизированной информационной системы «Профзаболевания» с целью оптимизации профилактических мероприятий. 2001г.-267 стр.

7. Дадонов Е.Т. и др. Автоматизированная система медико-санитарного обслуживания как подсистема промышленного комплекса. «Приборы и системы управления», 2005г.-345 стр.

8. Бессонов В.В. и др. Некоторые принципы и результаты разработки системы контроля производственных воздействий на среду и человека как подсистемы. «Проблемы контроля и защиты атмосферы от загрязнений». 2014г.-167 стр.

9. Единицы физических величин. ГОСТ8417-81.ССБТ-М. 1981 г.-211 стр.

10. Бурков В.Н., Грацианский Е.В., Дзюбко С.И., Щепин А.А. Методы и механизмы управления безопасностью М.СИНТЕГ, 2001г.-160 стр.