Проверка статистических гипотез

Содержание

Проверка гипотезы об однородности двух выборочных совокупностей

F-критерий Фишера

Критерий Фишера (Фишера – Снедекора) позволяет сравнивать генеральные дисперсии двух независимых выборок.

Нулевая гипотеза H_0 формулируется следующим образом: генеральные дисперсии двух выборок равны. Альтернативная гипотеза H_1: генеральные дисперсии двух выборок не равны.

АЛГОРИТМ

1. Сформулировать гипотезы. Выбрать уровень значимости α.
2. Найти эмпирическое значение критерия по формуле:

$$F_{эмп} = \frac{\sigma^2_{наибольшая}}{\sigma^2_{наименьшая}}. \qquad (1)$$

3. Найти число степеней свободы как $k_1 = n_1 - 1$ для выборки с наибольшей величиной дисперсии и $k_2 = n_2 - 1$ для выборки с наименьшей величиной дисперсии.
4. Определить критическое значение критерия Фишера по одноименной статистической таблице Приложения для степеней свободы k_1 (№ столбца таблицы), k_2 (№ строки таблицы) и уровня значимости $\alpha/2$.
5. Сравнить эмпирическое и критическое значения критерия Фишера, учитывая, что *F*-критерий правосторонний.

Если $F_{эмп} < F_{крит}$, то принимается нулевая гипотеза, иначе – альтернативная.

Пример. В двух третьих классах проводилось тестирование умственного развития по тесту ТУРМШ десяти учащихся. Полученные значения величин средних достоверно не различались, однако психолога интересует вопрос, есть ли различия в степени однородности показателей умственного развития между классами.

Решение. Нулевая гипотеза H_0: различия в степени однородности показателей умственного развития между классами статистически незначимы. Гипотеза H_1:

различия в степени однородности показателей умственного развития между классами статистически значимы. Возьмём уровень значимости α = 0,05. Результаты тестирования представлены в табл. 1.

Дисперсии для переменных X и Y равны: σ_x^2 = 572,83; σ_y^2 = 174,04.

По формуле (1) найдём эмпирическое значение F-критерия Фишера: $F_{эмп} = \frac{572{,}83}{174{,}04} = 3{,}29.$

Критическое значение F-критерия для степеней свободы $k_1 = k_2 = 10 - 1 = 9$ равно $F_{крит}(0{,}05/2;\ 9;\ 9) = 4{,}03$.

Таким образом, $F_{эмп} < F_{крит}$, так как 3,29 < 4,03, и на уровне значимости 0,05 принимаем нулевую гипотезу, то есть различия в степени однородности показателей умственного развития между классами статистически незначимы.

Таблица 1

№ учащихся	Первый класс	Второй класс
1	90	41
2	29	49
3	39	56
4	79	64
5	88	72
6	53	65
7	34	63
8	40	87
9	75	77
10	79	62
Суммы	606	636
Среднее	60,6	63,6

Пример. При изучении индивидуальной изменчивости озимой пшеницы было установлено, что высота растений определённого сорта на разных предшественниках равна: после кукурузы $\bar{x}_1$ = 108 см (выборка из 15 растений) и после гороха $\bar{x}_2$ = 111 см (выборка из 12 растений). Средние оценки дисперсий соответственно равны D_1 = 25 см^2 и D_2 = 16 см^2. Выяснить: выращенные на разных предшественниках растения действительно различаются по изменчивости высоты стебля или различие между средними оценками дисперсий является случайным, несущественным.

Решение. Сформулируем гипотезы. В качестве нулевой гипотезы возьмём гипотезу о равенстве генеральных дисперсий на уровне значимости α = 0,05.

Найдём эмпирическое значение критерия Фишера по формуле (1) имеем: $F_{эмп}$= 25/16 = 1,5625.

Определим для степеней свободы $k_1 = 15 - 1 = 14$, $k_2 = 12 - 1 = 11$ и уровня значимости 0,05 по статистической таблице критических значений критерия Фишера из Приложения $F_{крит}(0{,}05/2;\ 14;\ 11) = 3{,}36$.

Таким образом, $F_{эмп} < F_{крит}$, так как 1,5625 < 3,36, и на уровне значимости 0,05 принимается нулевая гипотеза, то есть различие изменчивости высоты стебля озимой пшеницы, выращенной на разных предшественниках, является случайным и несущественным.

Замечание. В некоторых случаях в качестве конкурирующей выдвигают гипотезу о превышении одной из генеральных дисперсий значения другой. Например, если оценка дисперсии S_1^2 оказалась большей, чем S_2^2, то в качестве конкурирующей выдвигают гипотезу $\sigma_1^2 > \sigma_2^2$. Тогда проверку нулевой гипотезы о равенстве генеральных дисперсий при уровне значимости α проводят тем же методом, что и при конкурирующей гипотезе $\sigma_1^2 \neq \sigma_2^2$, однако используют критическое значение распределения Фишера – Снедекора, соответствующее полной величине уровня значимости: $F_{крит}(\alpha;\ n_1 - 1;\ n_2 - 1)$.

Критерий Вилкоксона

Проверка гипотезы об однородности двух выборочных совокупностей может быть выполнена по критерию Вилкоксона. Критерий применим к случайным величинам, распределения которых неизвестны; требуется лишь, чтобы величины были непрерывными. Если выборки однородны, то считают, что они извлечены из одной и той же генеральной совокупности и, следовательно, имеют одинаковые, причем неизвестные, непрерывные функции распределения.

Пусть имеются две независимые выборки X и Y с объемами n и m соответственно из двух генеральных совокупностей с непрерывными функциями распределения равными соответственно F и G. Сформулируем гипотезы:

H_0: $F(x) = G(x)$ для всех $x \in R$.

H_1: 1) $F(x) \neq G(x)$; 2) $F(x) < G(x)$; 3) $F(x) > G(x)$; для всех $x \in R$.

Без ограничения общности будем считать, что $m \le n$. Составим объединенную выборку $Z = (X, Y)$. Построим вариационный ряд объединенной выборки: $z(1) < z(2) < \ldots < z(m+n)$.

Если распределения генеральных совокупностей непрерывны, то совпадения возможны только с нулевой вероятностью.

Найдем, какие места занимают в вариационном ряду, построенном по объединенной выборке, элементы выборки Y, назовем эти номера *рангами* элементов выборки Y в объединенной выборке Z:

$$rank(Y_1) = s_1,\ rank(Y_2) = s_2,\ \ldots,\ rank(Y_m) = s_m.$$

Рассмотрим статистку критерия:

$$W_{эмп} = \sum_{i=1}^{m} s_i\ .$$

Далее могут предложены разные подходы рассуждений.

1 подход. Очевидно, что минимальным значением статистики Вилкоксона может быть величина: $W_{min} = m\cdot(m+1)/2$, максимальным $W_{max} = mn + m\cdot(m+1)/2$.

Поэтому статистика $W_{эмп}$ находится в промежутке

$$[m\cdot(m+1)/2;\ m\cdot n + m\cdot(m+1)/2].$$

Распределение статистики Вилкоксона $W_{эмп}$ является симметричным относительно середины данного промежутка при условии справедливости нулевой гипотезы H_0.

Если справедлива альтернативная гипотеза H_1, то в случае: 2) распределение статистики W будет сдвинуто влево, так как чаще будут выполняться события $x_i > y_j$, 3) чаще будут встречаться события $x_i < y_j$, то есть, распределение статистики W перестанет быть симметричным относительно середины и будет сдвинуто вправо.

Если выбрана альтернативой гипотеза H_1, то критическая область для нулевой гипотезы H_0 будет иметь вид промежутка:

в случае 1) $[m\cdot(m+1)/2;\ c_2]$; $[c_1;\ m\cdot n + m\cdot(m+1)/2]$,

в случае 2) $[m\cdot(m+1)/2;\, c_2]$,
в случае 3) $[c_1;\, m\cdot n+m\cdot(m+1)/2]$.

При этом, константы c_1, c_2, c_3, c_4 следует находить по таблицам распределения статистики Вилкоксона W, рассчитанным при условии справедливости нулевой гипотезы H_0 для разных m и n.

В качестве искомых констант выбираются квантили распределения. При этом, константы c_1 и c_2 симметричны относительно середины промежутка $[m(m+1)/2;\, mn+m(m+1)/2]$. Также симметрично относительно середины этого промежутка расположены константы c_3 и c_4.

Общее требование заключается в том, что вероятность попадания статистики W в критическую область при условии справедливости нулевой гипотезы H_0 должна быть равна заданному значению α: $P_0\{W \in S\} = \alpha$.

2 подход. Находим (см. [3]):

1) при условии, что объемы выборок не больше 25, нижнюю границу по таблице значений Вилкоксона $W_{ниж.крит}(Q=\alpha/2;\, m;\, n)$; верхнюю границу по формуле: $W_{верх.крит} = (m+n+1)m - W_{ниж.крит}$.

2) при условии, что объемы выборок больше 25, нижнюю границу по таблице значений Вилкоксона

$$W_{ниж.крит}(Q=\alpha/2;\, m;\, n) = ((m+n+1)m-1)/2 - z_{крит}\sqrt{\frac{n\cdot m(n+m+1)}{12}},$$

где $z_{крит}$ с помощью MS Excel статистической функции: для случая 1) **НОРМ.СТ.ОБР((1- α)/2 + 0,5)**, для случаев 2), 3) **НОРМ.СТ.ОБР((1- 2α)/2 + 0,5)**; верхнюю границу по формуле: $W_{верх.крит} = (m+n+1)m - W_{ниж.крит}$..

Если $W_{эмп} \in [W_{ниж.крит};\, W_{верх.крит}]$, то принимается нулевая гипотеза, иначе альтернативная.

Пример. Проверить на однородность выборочные совокупности.

X	1	2	5	7	16	20	22
Y	3	4	6	10	13	17	

Решение. Сформулируем гипотезы:

H_0: $F(x) = G(x)$ (выборочные совокупности однородны).
H_1: $F(x) \neq G(x)$ (выборочные совокупности неоднородны).

Порядковый номер (ранг)	1	2	*3*	*4*	5	*6*	7	*8*	*9*	10	*11*	12	13
Варианта	1	2	*3*	*4*	5	*6*	7	*10*	*13*	16	*17*	20	22

Найдем, какие места занимают в вариационном ряду, построенном по объединенной выборке, элементы выборки Y, назовем эти номера *рангами* элементов выборки Y (столбцы с курсивом) и будем их складывать.

$W_{эмп}$ = 3+4+6+8+9+11=41.

1 подход. $W_{min} = m\cdot(m+1)/2$=6·(6+1)/2=21.

$W_{max} = mn + m\cdot(m + 1)/2$=6·7+6·(6+1)/2=63.

$W_{эмп}$ = 41, 41 ∈ [21; 63] и является симметричным относительно середины промежутка, поэтому можем сделать вывод о справедливости нулевой гипотезы H_0, то есть принимаем гипотезу об однородности выборочных совокупностей.

2 подход. Выберем уровень значимости 0,01, тогда найдем по таблице критических значений критерия Вилкоксона (см. ниже):

$W_{ниж.крит}$ (0,01/2; 6; 7) =24;

$W_{верх.крит} = (m + n+1)\cdot m - W_{ниж.крит}$ = (6+7+1)·6 – 24 = 60.

$W_{эмп}$ = 41, 41 ∈ [24; 60], то есть принимаем нулевую гипотезу об однородности выборочных совокупностей.

Критические точки критерия Вилкоксона

Объемы выборок		Q				Объемы выборок		Q			
n_1	n_2	0,005	0,01	0,025	0,05	n_1	n_2	0,005	0,01	0,025	0,05
6	6	23	24	26	28	7	7	32	34	36	39
	7	24	25	27	30		8	34	35	38	41
	8	25	27	29	31		9	35	37	40	43
	9	26	28	31	33		10	37	39	42	45
	10	27	29	32	35		11	38	40	44	47
	11	28	30	34	37		12	40	42	46	49
	12	30	32	35	38		13	41	44	48	52
	13	31	33	37	40		14	43	45	50	54
	14	32	34	38	42		15	44	47	52	56
	15	33	36	40	44		16	46	49	54	58
	16	34	37	42	46		17	47	51	56	61
	17	36	39	43	47		18	49	52	58	63
	18	37	40	45	49		19	50	54	60	65
	19	38	41	46	51		20	52	56	62	67
	20	39	43	48	53		21	53	58	64	69
	21	40	44	50	55		22	55	59	66	72
	22	42	45	51	57		23	57	61	68	74
	23	43	47	53	58		24	58	63	70	76
	24	44	48	54	60		25	60	64	72	78
	25	45	50	56	62						

Объемы выборок		Q			
n_1	n_2	0.005	0.01	0.025	0.05
8	8	43	45	49	51
	9	45	47	51	54
	10	47	49	53	56
	11	49	51	55	59
	12	51	53	58	62
	13	53	56	60	64
	14	54	58	62	67
	15	56	60	65	69
	16	58	62	67	72
	17	60	64	70	75
	18	62	66	72	77
	19	64	68	74	80
	20	66	70	77	83
	21	68	72	79	85
	22	70	74	81	88
	23	71	76	84	90
	24	73	78	86	93
	25	75	81	89	96
9	9	56	59	62	66
	10	58	61	65	69
	11	61	63	68	72
	12	63	66	71	75
	13	65	68	73	78
	14	67	71	76	81
	15	69	73	79	84
	16	72	76	82	87
	17	74	78	84	90
	18	76	81	87	93
	19	78	83	90	96
	20	81	85	93	99
	21	83	88	95	102
	22	85	90	98	105
	23	88	93	101	108
	24	90	95	104	111
	25	92	98	107	114
10	10	71	74	78	82
	11	73	77	81	86
	12	76	79	84	89
	13	79	82	88	92
	14	81	85	91	96
	15	84	88	94	99
	16	86	91	97	103
	17	89	93	100	106
	18	92	96	103	110
	19	94	99	107	113
	20	97	102	110	117
	21	99	105	113	120
	22	102	108	116	123
	23	105	110	119	127
	24	107	113	122	130
	25	110	116	126	134
11	11	87	91	96	100
	12	90	94	99	104
	13	93	97	103	108
	14	96	100	106	112
	15	99	103	110	116
	16	102	107	113	120
	17	105	110	117	123
	18	108	113	121	127
	19	111	116	124	131
	20	114	119	128	135
	21	117	123	131	139
	22	120	126	135	143
	23	123	129	139	147
	24	126	132	142	151
	25	129	136	146	155
12	12	105	109	115	120
	13	109	113	119	125
	14	112	116	123	129
	15	115	120	127	133
	16	119	124	131	138
	17	122	127	135	142
	18	125	131	139	146
	19	129	134	143	150
	20	132	138	147	155
	21	136	142	151	159
	22	139	145	155	163
	23	142	149	159	168
	24	146	153	163	172
	25	149	156	167	176
13	13	125	130	136	142
	14	129	134	141	147
	15	133	138	145	152
	16	136	142	150	156
	17	140	146	154	161
	18	144	150	158	166

Сравнение двух средних генеральных совокупностей

1. Сравнение двух средних генеральных совокупностей, дисперсии которых известны.

Рассмотрим статистический критерий, который используется для сравнения генеральных средних двух больших независимых выборок, случайным образом отобранных из двух нормально распределенных генеральных совокупностей *X* и *Y*, чьи дисперсии известны (например, из предшествующего опыта или найдены теоретически), в некоторых источниках его называют критерий Лапласа.

Гипотезы. Нулевая гипотеза H_0: генеральные средние равны $\overline{X} = \overline{Y}$, тогда H_1: ***первый случай***: генеральные средние различны; ***второй случай***: $\overline{X} > \overline{Y}$; ***третий случай***: $\overline{X} < \overline{Y}$.

АЛГОРИТМ

1. Сформулировать гипотезы. Выбрать уровень значимости α.

2. Найти эмпирическое значение критерия Лапласа по формуле: $z_{эмп} = \dfrac{\left|\bar{x} - \bar{y}\right|}{\sqrt{\dfrac{D_x}{n_x} + \dfrac{D_y}{n_y}}}$,

где $\bar{x}$ и $\bar{y}$ – выборочные средние, D_1 и D_2 – известные дисперсии генеральных совокупностей.

3. Найти критическое значение критерия по статистической одноимённой таблице Приложения: в первом случае: $\Phi\left(z_{крит}\right) = \dfrac{p}{2} = \dfrac{1-\alpha}{2}$, во втором и третьем случаях: $\Phi\left(z_{крит}\right) = \dfrac{1-2\alpha}{2}$, где $\Phi(L)$ – интегральная функция Лапласа.

4. Сравнить эмпирическое и критическое значения критерия.

Если $z_{эмп} < z_{крит}$, то на уровне значимости α принимается нулевая гипотеза H_0, иначе – альтернативная.

Пример. Необходимо исследовать на возможную высокую урожайность два сорта пшеницы. Для этого составляют две выборки: равномерно и случайным образом со всей площади посева каждого изучаемого сорта берут на исследование по 50 или более растений. Пусть для анализа сорта пшеницы выбран хозяйственно важный признак – число колосков в колосе. Пусть собрано по 100 колосьев со 100 разных растений каждого сорта, далее подсчитывается число колосков (X для первого и Y для второго сортов, их объёмы n_x и n_y) в каждом колосе и составляются два вариационных ряда для первого и второго сортов. Выполнить с надежностью 99%.

X	n_x
12	3
13	8
14	16
15	19
16	21
17	17
18	9
19	5
20	2

Y	n_y
11	1
12	4
13	4
14	7
15	12
16	18
17	22
18	15
19	8
20	6
21	3

Решение. Сформулируем гипотезы. Нулевая гипотеза H_0: существенных различий между исследуемыми сортами нет. H_1: второй сорт существенно превосходит первый по изучаемому признаку (третий случай).

Для более полного представления об изменчивости изучаемых сортов пшеницы по числу колосков вычислим среднее арифметическое ($\bar{x}$, $\bar{y}$ для первого и второго сортов соответственно) и соответствующие им дисперсии:

$$\bar{x} = \sum_{i=1}^{9} x_i \cdot n_{xi} = 15{,}71 \text{ и } \bar{y} = \sum_{i=1}^{11} y_i \cdot n_{yi} = 16{,}56.$$

$$D_x = \frac{\sum_{i=1}^{9}(x_i - 15,71)^2 \cdot n_{xi}}{100} = 3,24 \text{ и } D_y = \frac{\sum_{i=1}^{11}(y_i - 16,56)^2 \cdot n_{yi}}{100} = 4,58.$$

Определим существенность различия сравниваемых величин, то есть существенность различий двух сортов пшеницы по числу колосков в колосе.

$$z_{эмп} = \frac{\left|\bar{x} - \bar{y}\right|}{\sqrt{\frac{D_x}{n_x} + \frac{D_y}{n_y}}} = \frac{\left|15,71 - 16,56\right|}{0,1 \cdot \sqrt{3,24 + 4,58}} \approx 3,04.$$

Для уровня значимости α = 0,01 определим критическое значение критерия по статистической таблице функции Лапласа (третий случай).

$$\Phi\left(z_{крит}\right) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0,01}{2} = 0,49 \text{ и } z_{\text{крит}} = 2{,}33.$$

Таким образом, $z_{эмп} > z_{крит}$, так как 3,04 > 2,33, и на уровне значимости 0,01 принимаем альтернативную гипотезу, то есть второй сорт существенно превосходит первый по изучаемому признаку, при этом различие количества колосков имеет место не только в данных выборках, но и в генеральных совокупностях в целом (с надёжностью 0,99).

Если учесть, что увеличение числа колосков в колосе, как правило, сопровождается увеличением числа зёрен, то при прочих равных характеристиках от этого сорта мы вправе ожидать более высокой урожайности по сравнению с другим сортом.

2. Сравнение двух средних генеральных совокупностей, дисперсии которых неизвестны.

Рассмотрим статистический критерий, который используется для сравнения генеральных средних двух нормально распределённых случайных величин при неизвестных дисперсиях генеральных совокупностей.

При использовании критерия (t-критерий Стьюедента) можно выделить два случая. В первом случае критерий применяется для двух **независимых**, **несвязанных** выборок (двухвыборочный t-критерий), во втором случае – для **зависимых**, **связанных** выборок (парный t-критерий).

Отметим, что в обоих случаях гипотезы будут формулироваться следующим образом.

Гипотезы. Нулевая гипотеза H_0: генеральные средние значения равны, альтернативная гипотеза H_1: генеральные средние значения различны.

Первый случай (случай независимых выборок). В случае независимых выборок t-критерий применим для малых независимых выборок и если есть основания считать генеральные дисперсии равными.

АЛГОРИТМ

1. Сформулировать гипотезы. Выбрать уровень значимости α.

2. Найти эмпирическое значение критерия Стьюдента для случая несвязанных, независимых выборок по формуле:

$$t_{эмп}=\frac{\bar{x}-\bar{y}}{\sigma_{x-y}}=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2}{n_x+n_y-2}\cdot\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}},$$

где $\bar{x}, \bar{y}, \sigma_{x-y}$ – это средние арифметические в экспериментальной и контрольной группах, стандартная ошибка разности средних арифметических соответственно.

Замечание. В некоторых источниках вышеприведённая формула имеет вид:

$$t_{эмп}=\frac{|\bar{x}-\bar{y}|}{\sqrt{(n_x-1)\cdot S_x^2+(n_y-1)\cdot S_y^2}}\sqrt{\frac{n_x n_y(n_x+n_y-2)}{n_x+n_y}}, \qquad (33)$$

где $\bar{x}, \bar{y}$; S_x^2, S_y^2; n_x и n_y – средние арифметические, «исправленные» дисперсии, объёмы первой и второй выборок соответственно. В ряде источников этот критерий носит название Стьюдента – Фишера.

4. Найти число степеней свободы по формуле: $k = n_x + n_y - 2$.

5. Определить критическое значение t-критерия Стьюдента по статистической одноименной таблице из Приложения.

6. Сравнить эмпирическое и критическое значения критерия Стьюдента, учитывая, что t-критерий двусторонний.

Если $|t_{эмп}| < |t_{крит}|$, принимается нулевая гипотеза, иначе принимается альтернативная.

Замечание. В ряде научных исследований применяется «приближение» t-критерия:

$$t_{эмп}=\frac{|\bar{x}-\bar{y}|}{\sqrt{n_x\cdot D_x+n_y\cdot D_y}}\sqrt{n_x n_y}. \qquad (34)$$

В ряде источников это «приближение» называют критерием Крамера – Уэлча. Заметим, что критические значения для критерия Крамера – Уэлча зависят только от уровня значимости α и выражаются через критические значения t-критерия Стьюдента следующим образом: $t_{крит}(0,01; \infty) = 2,58$, $t_{крит}(0,1; \infty) = 1,65$, $t_{крит}(0,05; \infty) = 1,96$.

Алгоритм и схема использования критерия Крамера – Уэлча аналогичные, как и для критерия Стьюдента. При этом очевидно, что чем больше объёмы выборок, тем меньше отличия числовых значений данных критериев.

Пример. Измерено 10 колосьев суперэлиты и 12 колосьев элиты озимой ржи определённого сорта. Для каждой репродукции была вычислена средняя длина колосьев и средние оценки дисперсий. Для суперэлиты они соответственно равны $\bar{x} = 10,24$ см и $S_x^2 = 0,0169$ см^2, для элиты – $\bar{y} = 9,69$ см и $S_y^2 = 0,0256$ см^2. Выяснить, действительно ли различия в длинах колосьев обусловлены принадлежностью к элите (суперэлите) или же это различие является случайным, не существенным.

Решение. Сформулируем гипотезы. В качестве нулевой принимаем гипотезу: различия в средних длинах колосьев суперэлиты и элиты обусловлены случайностями выборки.

Определим эмпирическое значение критерия $t_{эмп}$ по формуле (33):

$$t_{эмп} = \frac{|\bar{x} - \bar{y}|}{\sqrt{(n_x - 1)\cdot S_x^2 + (n_y - 1)\cdot S_y^2}} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} =$$

$$= \frac{|10{,}24 - 9{,}69|}{\sqrt{(10-1)\cdot 0{,}0169 + (12-1)\cdot 0{,}0256}} \cdot \sqrt{\frac{10\cdot 12\cdot (10+12-2)}{10+12}} = \frac{0{,}55}{\sqrt{0{,}4337}} \cdot \sqrt{\frac{2400}{22}} \approx 8{,}72.$$

Найдём для уровня значимости 0,05 и степени свободы k = 10 + 12 – 2 = 20 критическое значение критерия $t_{крит}$(0,05; 20) = 2,09.

Таким образом, $t_{эмп}$ > $t_{крит}$, так как 8,72 > 2,09, и на уровне значимости 0,05 принимаем альтернативную гипотезу, то есть различия в средних длинах колосьев суперэлиты и элиты значимы и не обусловлены случайными причинами.

Пример. В двух группах учащихся – экспериментальной и контрольной – получены следующие результаты по учебному предмету (тестовые баллы).

Группа 1 (экспериментальная)	Группа 2 (контрольная)
12 14 13 16 11 9 13 15 15 18 14	13 9 11 10 7 6 8 10 11

Решение. Нулевая гипотеза: знания учащихся в обеих группах в среднем одинаковые, альтернативная гипотеза: знания учащихся экспериментальной группы отличаются от знаний учащихся контрольной группы.

Здесь из таблицы данных получим: n_x = 11, n_y = 9, $\bar{x}$ = 13,636, $\bar{y}$ = 9,444, σ_x = 2,460 и σ_y = 2,186.

Применим к данным задачи критерий Стьюдента.

Рассчитаем стандартную ошибку разности арифметических средних:

$$\sigma_{x-y} = \sqrt{\frac{60{,}545 + 38{,}222}{11 + 9 - 2} \cdot \left(\frac{1}{11} + \frac{1}{9}\right)} = 1{,}053.$$

Эмпирическое значение критерия $t_{эмп}$ равно:

$$t_{эмп} = \frac{13{,}636 - 9{,}444}{1{,}053} = 3{,}981.$$

Найдём для уровня значимости 0,05 и степени свободы k = 11 + 9 – 2 = 18 критическое значение критерия $t_{крит}$(0,05; 18) = 2,1.

Таким образом, $t_{эмп}$ > $t_{крит}$, так как 3,981 > 2,1, и на уровне значимости 0,05 принимается альтернативная гипотеза H_1, то есть можем утверждать о преимуществе экспериментального обучения.

Теперь к условию этой же задачи применим критерий Крамера – Уэлча.

Найдём эмпирическое значение критерия $t_{эмп}$ по формуле (34):

$$t_{эмп} = \frac{|\bar{x} - \bar{y}|}{\sqrt{n_x \cdot D_x + n_y \cdot D_y}} \sqrt{n_x n_y} = \frac{|13{,}636 - 9{,}444|}{\sqrt{11\cdot 2{,}46^2 + 9\cdot 2{,}186^2}} \sqrt{11\cdot 9} = 3{,}985.$$

Критическое значение критерия для уровня значимости 0,05 равно $t_{крит}$ = 1,96.

Таким образом, $t_{эмп} > t_{крит}$, так как 3,985 > 1,96, и на уровне значимости 0,05 принимается альтернативная гипотеза H_1, то есть преимущество экспериментального обучения статистически значимо.

Замечание. Обратим внимание, что эмпирические значения, найденные по формулам критериев Стьюдента и его аналога – критерия Крамера – Уэлча, мало отличимы (третьей значащей цифрой после запятой).

Второй случай (случай зависимых, связанных выборок).

АЛГОРИТМ

1. Сформулировать гипотезы. Выбрать уровень значимости α.

2. Найти эмпирическое значение критерия Стьюдента для случая связанных выборок с равным числом измерений по формуле:

$$t_{эмп} = \frac{\bar{d}}{S_d}, \tag{35}$$

где $d_i = x_i - y_i$ – разности между соответствующими значениями переменной X и переменной Y, а $\bar{d}$ – среднее этих разностей.

3. Найти S_d по формуле:

$$S_d = \sqrt{\frac{\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n \cdot (n-1)}}.$$

4. Найти число степеней свободы k по формуле $k = n - 1$.

5. Определить критическое значение критерия Стьюдента по статистической одноимённой таблице Приложения.

6. Сравнить эмпирическое и критическое значения критерия Стьюдента, учитывая, что t-критерий двусторонний.

Если $|t_{эмп}| < |t_{крит}|$, то принимается нулевая гипотеза, иначе принимается альтернативная.

Рассмотрим пример использования t-критерия Стьюдента для связных и, очевидно, равных по численности выборок.

Пример. Изучалось влияние курения на агрегацию эритроцитов, у 10 мужчин-добровольцев агрегация (мм/ч) измерялась до и после выкуривания одной сигареты.

До	После	D	d^2
2	2	0	0
2	9	7	49
5	11	6	36
7	11	4	16
8	14	6	36
9	9	0	0
10	9	-1	1
10	10	0	0
10	13	3	9
10	15	5	25
Сумма		30	172

Решение. Сформулируем гипотезы. Нулевая гипотеза H_0: выкуривание одной сигареты на агрегацию эритроцитов влияет несущественно, H_1: выкуривание одной сигареты на агрегацию эритроцитов влияет существенно.

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \frac{30}{10} = 3. \quad S_d = \sqrt{\frac{\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n \cdot (n-1)}} = \sqrt{\frac{172 - (30)^2/10}{10 \cdot (10-1)}} = \sqrt{0{,}91} \approx 0{,}955.$$

По формуле (35) найдём $t_{эмп} = \frac{\bar{d}}{S_d} = \frac{3}{0{,}955} = 3{,}14$. По статистической таблице определим критическое значение критерия для числа степеней свободы $k = 10 - 1 = 9$ на уровне значимости 0,01. $t_{крит} = 3{,}25$.

Таким образом, $t_{эмп} < t_{крит}$, так как 3,14 < 3,25, и на уровне значимости 0,01 принимается гипотеза H_0, то есть изменения агрегации эритроцитов при выкуривании одной сигареты несущественны.

Замечание. Критерий Стьюдента можно применять для исключения грубых ошибок наблюдений, возникающих из-за показаний измерительных приборов, ошибок регистрации, случайного сдвига запятой в десятичной записи числа и т. п..

Пусть, например, $x^*, x_1, x_2, ..., x_n$ – совокупность статистических данных, из которых первое значение резко выделяется из других.

Необходимо выяснить, принадлежит ли резко выделяющееся значение к остальным наблюдениям.

Для ряда $x_1, x_2, ..., x_n$ рассчитывают среднее арифметическое $\bar{x}$ и «исправленное» среднее квадратическое отклонение s.

Нулевая гипотеза H_0: о принадлежности x^* к остальным наблюдениям. H_1: о не принадлежности x^* к остальным наблюдениям.

Эмпирическое значение критерия Стьюдента равно: $t_{эмп} = \left|\frac{\bar{x} - x^*}{s}\right|$.

Критическое значение критерия Стьюдента для уровня значимости α и степени свободы $k = n - 2$ находят по статистической таблице Стьюдента из Приложения.

Если $t_{эмп} < t_{крит}$, то принимают нулевую гипотезу, иначе – альтернативную.

Пример. Имеются следующие данные об урожайности ржи (ц/га) на 8 опытных участках одинаковой площади: 25; 26,1; 26,2; 26,5; 29,3; 30,1; 32,3; 35,9. Выяснить, является ли значение урожайности $x^* = 35{,}9$ аномальным (неверно зарегистрировано) при уровне значимости 0,05 [48].

Решение. Сформулируем гипотезы. Нулевая гипотеза H_0: о принадлежности x^* к остальным наблюдениям. H_1: о непринадлежности x^* к остальным наблюдениям.

Исключим значение $x^* = 35{,}9$ из выборки. Для новой выборки:

25; 26,1; 26,2; 26,5; 29,3; 30,1; 32,3

найдём среднее значение $\bar{x}$ = 27,93 (ц/га), «исправленное» среднее квадратическое отклонение s = 2,67 (ц/га).

Эмпирическое значение критерия Стьюдента равно $t_{эмп} = \left|\frac{\bar{x} - x^*}{s}\right| = \left|\frac{35{,}9 - 27{,}93}{2{,}67}\right| = 2{,}98$.

Критическое значение критерия Стьюдента равно $t_{крит} = t(0{,}05;\, 6) = 2{,}45$.

Таким образом, $t_{эмп} > t_{крит}$, так как 2,98 > 2,45, и на уровне значимости 0,05 принимается альтернативная гипотеза, то есть значение x^* = 35,9 является аномальным и его следует отбросить.

Замечание. Критерий Стьюдента предназначен для сравнения двух выборок. Между тем на практике его порой неправильно используют для оценки различий большего числа групп посредством попарного их сравнения. При этом вступает в силу ***эффект множественных сравнений***. Вероятность ошибиться хотя бы в одном из нескольких сравнений составит: $\alpha' = 1 - (1 - 0{,}01)^k$, где k – число сравнений (если надёжность берём 0,99). При небольшом числе сравнений можно использовать приближённую формулу: $\alpha' = 0{,}01 \cdot k$.

Точные значения вероятности ошибки при различном числе групп приведены в следующей таблице:

Количество групп	Количество попарных сравнений	α' (α = 0,01)	α' (α = 0,05)
3	3	0,029	0,143
4	6	0,059	0,265
5	10	0,096	0,401

Таким образом, если мы проводим попарные сравнения нескольких групп и хотим с вероятностью ошибки 0,01 найти различия, необходимо полученную после применения критерия Стьюдента точную вероятность α умножить приближённо на количество попарных сравнений. Например, мы проводим попарные сравнения трёх групп, таких сравнений будет три: 1–2, 1–3, 2–3. Тогда, чтобы отклонить нулевую гипотезу об отсутствии различий между сравниваемыми группами на уровне значимости 0,01, вероятность ошибки должна составить не более 0,01/3 = 0,0033. Можно также сравнивать эмпирическое значение с критическим значением, соответствующим этому значению вероятности, которое можно приближённо рассчитать по формуле: $t_н = t_1 + (t_2 - t_1)\dfrac{\alpha_н - \alpha_1}{\alpha_2 - \alpha_1}$, где α_1, α_2 – значения уровня значимости; t_1, t_2 – критические значения, соответствующие им; $\alpha_н$ – нужный уровень значимости.

Этот же подход можно использовать при применении других критериев, предназначенных для сравнения двух групп, для множественных сравнений.

Проверка гипотезы о незначимости коэффициента корреляции

Оценим достоверность коэффициента корреляции по формуле:

$$t_r = \frac{|r_{xy}|}{S_r} = |r_{xy}| \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}},$$

Найдем критическое значение критерия для уровня значимости α и для числа степеней свободы $k = n - 2$ с помощью статистической функции **СТЬЮДЕНТ.ОБР.2Х(*α*; *k*)**.

Вывод:

1) если $t_r \geq t_{крит}$, то отвергают нулевую гипотезу H_0, то есть при выбранном уровне значимости делают вывод о статистической значимости коэффициента корреляции;

2) если $t_r < t_{крит}$, принимают нулевую гипотезу H_0, то есть при выбранном уровне значимости делают вывод о статистической значимости коэффициента корреляции.

В MS Excel для вычисления парных коэффициентов линейной корреляции используется специальная функция **КОРРЕЛ (массив1; массив2),** где

массив1 – ссылка на диапазон ячеек первой выборки (X);

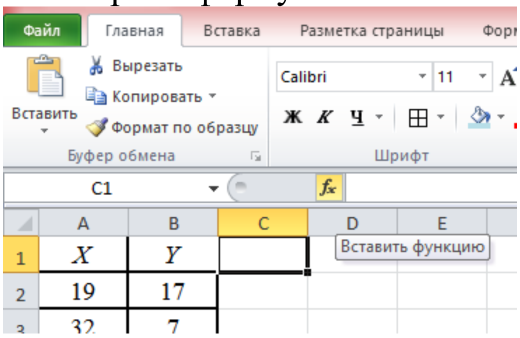
массив2 – ссылка на диапазон ячеек второй выборки (Y).

Пример. Десяти школьникам были даны тесты на наглядно-образное и вербальное мышление. Измерялось среднее время решения заданий теста в секундах. Исследователя интересует вопрос: существует ли взаимосвязь между временем решения этих задач? Переменная X обозначает среднее время решения наглядно-образных, а переменная Y – среднее время решения вербальных заданий тестов (табл. 2).

Таблица 2

№	1	2	3	4	5	6	7	8	9	10
X	19	32	33	44	28	35	39	39	44	44
Y	17	7	17	28	27	31	20	17	35	43

Решение. Введем данные в таблицу MS Excel. Затем вычислим значение коэффициента корреляции. Для этого курсор установим в ячейку С1 и активизируем кнопку f_x, находящуюся слева от строки формул.



В появившемся диалоговом окне выберем функцию **КОРРЕЛ** категории **Статистические**. Указателем мыши введем диапазон данных выборки X в поле *массив1* (A1:A10). В поле *массив2* введем диапазон данных выборки Y (B1:B10), нажмем кнопку **ОК**. В ячейке С1 появится значение коэффициента корреляции 0,54119.

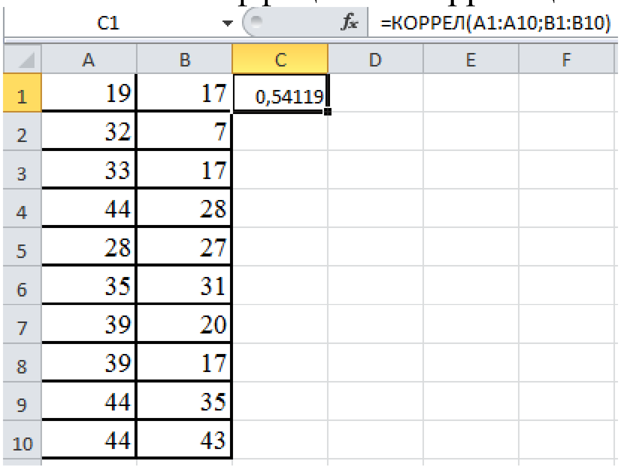


Рис. 1. Результаты вычисления коэффициента корреляции

Далее оценим достоверность коэффициента корреляции по формуле:

$$t_r = \left| r_{xy} \right| \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}} = 0{,}54 \cdot \sqrt{\frac{10-2}{1-0{,}54^2}} = 1{,}815.$$

Найдем критическое значение критерия с помощью статистической функции **СТЬЮДЕНТ.ОБР.2Х(0,05; 8)**: $t_{крит} = 2{,}306$.

Таким образом, $t_r < t_{крит}$, так как 1,815 < 2,306, и на уровне значимости 0,05 принимается гипотеза H_0, иными словами, связь между временем решения наглядно-образных и вербальных заданий теста не доказана.

Литература

1. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли. – М.: Прогресс, 1976. – 496 с.
2. Гланц С. Медико-биологическая статистика / С. Гланц. - М.: Практика, 1998. - 459 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – М.: Высш. шк., 2003. – 479 с.
4. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи) / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 67 с.
5. Новиков Д. А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д. А. Новиков, В. В. Новочадов. – Волгоград: Изд-во ВГМУ, 2005. – 84 с.
6. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Киров: Изд-во ВГГУ, 2015. – 158 с.