



Вторичная математическая обработка данных

Выбор статистического метода и классификация исследовательских задач

В целом в студенческих работах можно выделить несколько групп задач, которые наиболее часто решаются в экспериментальных исследованиях:

1. Выявление сходства или различия в уровне исследуемого признака.
2. Оценка сдвига значений исследуемого признака.
3. Выявление степени согласованности изменений признаков.

Классификация задач и методов их решения

№	Задачи	Условия	Метод (критерий) математической обработки
1.	Выявление различий между выборками по уровню исследуемого признака	а) две выборки испытуемых; желательны $n > 10$ наблюдений;	Параметрический: t – Стьюдента ;
		< 11 наблюдений;	Непараметрический: U – Манна – Уитни
2.	Оценка достоверности изменений исследуемого признака (сдвига значений)	а) оценка сдвига значений в экспериментальной выборке;	Непараметрический: G – критерий знаков
		б) сопоставление сдвигов в двух выборках, напр., в контрольной и экспериментальной группах;	Непараметрические критерии: U – Манна – Уитни.
		в) независимая обработка двух выборок;	G – критерий знаков
3.	Установления взаимосвязи (степени согласованности изменений) между уровнем выраженности исследуемого признака	а) данные измерены в шкале наименований : каждый признак должен быть представлен двумя альтернативными значениями признака;	Коэффициенты корреляции: Q - Юла и ассоциации - Ф ; φ - Пирсона для дихотомических данных
		б) данные измерены в шкале порядка : необходимо два ряда значений, которые могут быть проранжированы.	Коэффициенты ранговой корреляции: r_s - Спирмена и T - Кендалла для связанных рангов

Рекомендации к выбору критерия.

- Необходимо определить, является ли выборка зависимой или независимой.
- Следует знать объём каждой выборки, её однородность – неоднородность, нормальность распределения признака в ней.
- При выборе критерия необходимо знать его ограничения. Начинать обработку лучше с наименее трудоёмкого метода; если он не позволил установить различия, следует применить другой – более мощный.
- Следует помнить, что при малых объёмах выборки многие критерии становятся чувствительными к форме распределения признака и при этом необходимо увеличивать величину уровня значимости не менее 1 %, чтобы не принять ошибочное решение.

Выявление различий между выборками.

Параметрические статистические критерии. t – Стьюдента (критерий Госсета).

- Основными параметрами критерия являются дисперсия (s^2) и среднее арифметическое значение (M).
- Критерий даёт информацию о том, достоверно ли различие средних значений двух выборок, но данные должны соответствовать *закону нормального распределения*.
- Если объём выборки меньше 10 наблюдений, лучше прибегнуть к непараметрическим критериям.

- После его применения можно говорить о том:
 1. удался эксперимент, оказал ли он влияние на уровень развития изучаемого признака (при этом замеры признака в одной и той же выборке делают в начале и в конце экспериментального исследования)
 2. различаются ли группа А и группа В по выраженности изучаемого признака.

Параметрические статистические критерии. *t* – Стьюдента (критерий Госсета).

Алгоритм нахождения *t* критерия:

1. Находим S_1^2 и S_2^2 ;
2. Находим величины их статистических ошибок по формуле:

$$m_1^2 = \frac{S_1^2}{n_1} \qquad m_2^2 = \frac{S_2^2}{n_2}$$

3. Находим формуле: $t_{эмп}$ (эмпирическое значение *t* – критерия) по

$$t = \frac{|M_1 - M_2|}{\sqrt{m_1^2 + m_2^2}}$$

4. Находим (критическое значение) по таблице для 95% и 99% уровня значимости. $p < 0,05$ и $p < 0,01$, для этого находим число степеней свободы по формуле:
где n – объёмы сравниваемых выборок. $d = n_1 + n_2 - 2$,

Параметрические статистические критерии. t – Стьюдента (критерий Госсета).

Сравниваем $t_{эмп}$ и $t_{кр}$, делаем выводы, если эмпирическое значение превышает критические (табличные) для $p < 0,01$, то различия статистически достоверны.

Таблица критических значений t – критерия
Стьюдента

Число степеней свободы к	р			Число степеней свободы к	р		
	0,05	0,01	0,001		0,05	0,01	0,001
1	12,71	63,66	64,60	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,08	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	∞	1,96	2,58	3,29

Непараметрические критерий **U – Манна – Уитни**

1. предназначен для оценки различий между двумя выборками по уровню выраженности признака, измеренного количественно.
2. применять можно на малых выборках при $n = 3$ и более наблюдений.
3. Данный критерий предполагает ранжирование переменных на основе определённых правил.

Алгоритм нахождения U критерия:

1. Помечаем показатели первой выборки одним цветом, а индивидуальные показатели второй выборки другим цветом.
2. Записываем данные в один ряд по нарастанию признака независимо от того, к какой выборке они принадлежат.
3. Ранжируем значения. Всего рангов будет $n_1 + n_2$
4. Считаем сумму рангов, ориентируясь при этом на цвет показателей. Проверяем, совпадает ли она с расчётной суммой.
5. Считаем сумму рангов, ориентируясь при этом на цвет показателей. Проверяем, совпадает ли она с расчётной суммой.
6. Определяем большую ранговую сумму.
7. Определяем Uэмп по формуле:

где, n_1 и n_2 - число наблюдений, объёмы выборок;
- количество испытуемых с в группе с большей суммой рангов;
 T_x – большая ранговая сумма.

$$U = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x$$

Алгоритм нахождения U критерия:

- Определяем критические значения по таблице, если $U_{эмп} < U_{кр}$ для $p < 0,05$, то различия достоверны/
- **Примечание:** в случаях, если объём выборок не совпадает, то необходимо рассчитать $U_{эмп}$ и для меньшей ранговой суммы;
- критерий U – Манна – Уитни является обратным, и поэтому чем меньше значение $U_{эмп}$, тем достовернее различия!



Проведено исследование вербального интеллекта в подгруппе студентов **первого курса (n=10)** и второго курса (n=11). Необходимо установить, различаются ли группы между собой по выраженности рассматриваемого признака.

задача: необходимо установить связь показателей интеллекта, полученных на одной выборке разными методами.

<i>127</i>
<i>107</i>
<i>115</i>
<i>116</i>
<i>111</i>
<i>115</i>
<i>107</i>
<i>104</i>
<i>106</i>
<i>107</i>

123
114
117
120
114
110
110
108
108
100