

Школа молодого исследователя

Лекция

Тема: «Алгоритм построения и проверки научной гипотезы»

Важность построения гипотезы

«Всякая плодотворная гипотеза
кладёт начало удивительному
извержению потока непредвиденных
открытий»



Леон Николя Бриллюэн
французский и американский физик

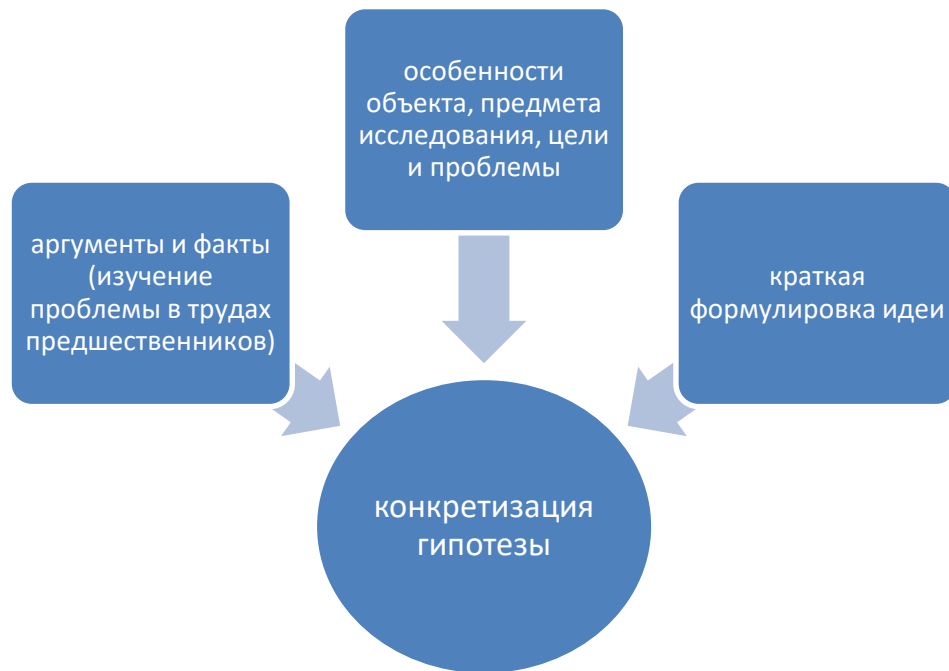
Принципы логического построения научного знания

1. Состоятельность выдвигаемой гипотезы
2. Применимость создаваемого знания
3. Проверяемость

Требования предъявляемые к гипотезе

- Совместимость с существующими знаниями
- Понятность и логичность
- Обоснованность (релевантность)
- Она должна быть проверяемой

Конкретизация гипотезы



Основные этапы проверки гипотезы



Гипотеза

1.

- Предположение о связи между двумя переменными: причина, следствие

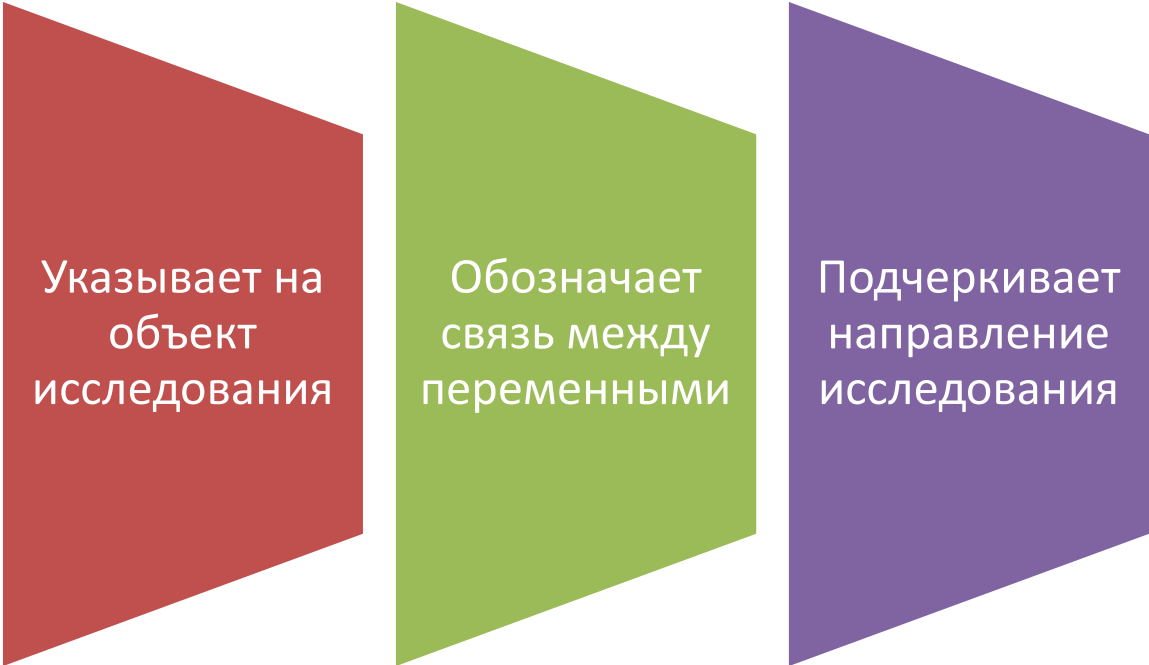
2.

- Предположение о сходстве и различии сравниваемых параметров или объектов

3.

- Принадлежность определенных характеристик объекту исследования

О чем говорит гипотеза



Указывает на
объект
исследования

Обозначает
связь между
переменными

Подчеркивает
направление
исследования

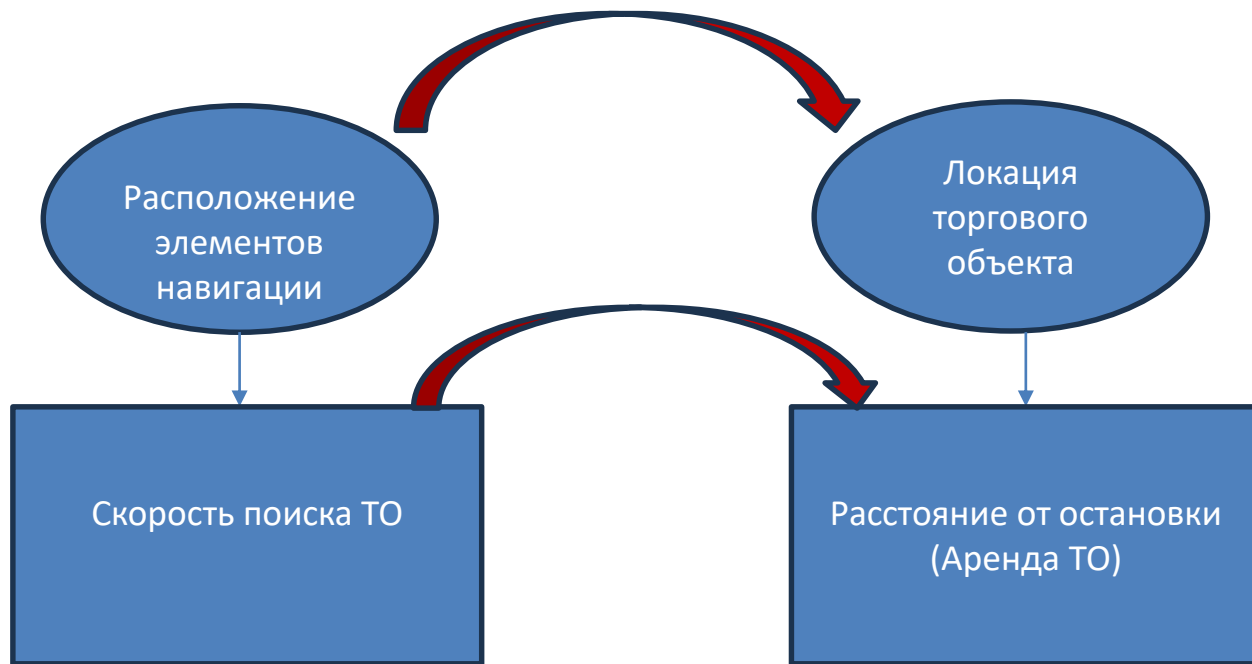
Примеры научных гипотез

Гипотеза 1. Эффективное расположение элементов навигации позволит повысить скорость поиска розничного объекта, уменьшив временные затраты клиента, что отразится на возможности выбора местоположения торговой точки в малопривлекательных городских локациях с меньшей стоимостью арендной ставки.

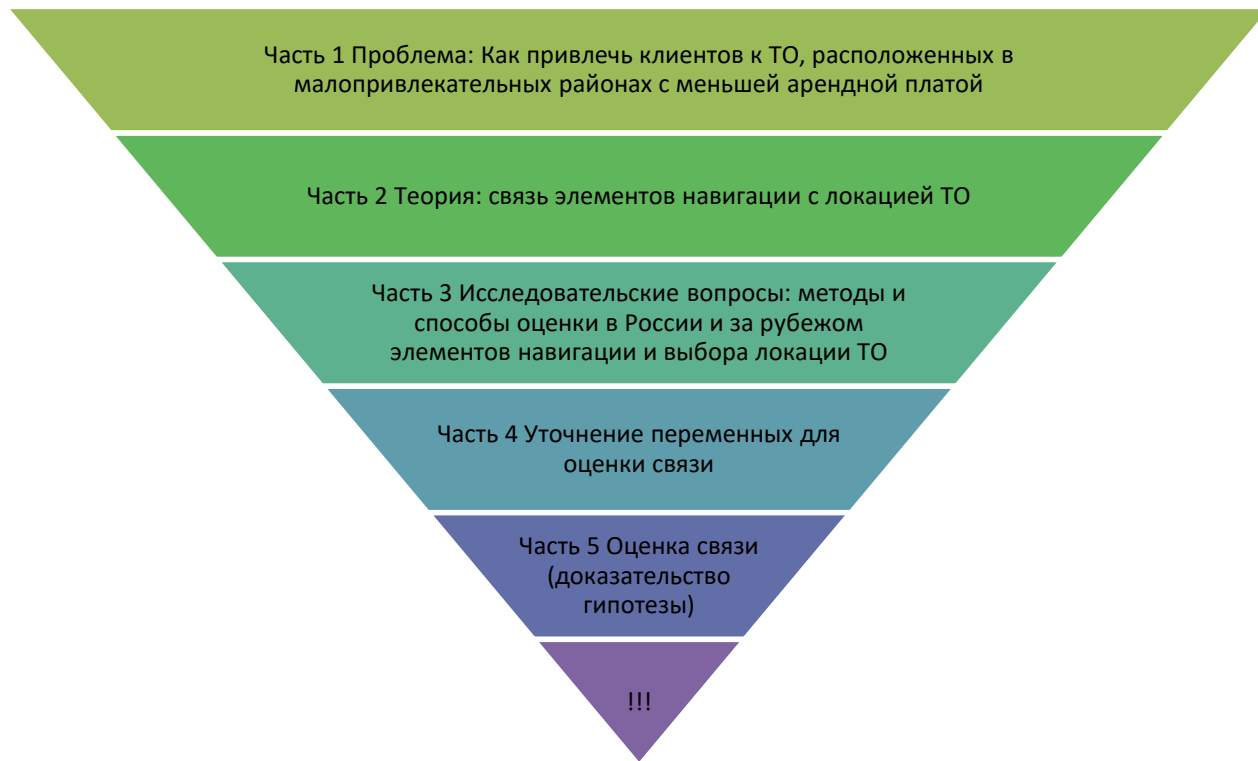
Гипотеза 2. Пространственная дифференциация регионов по уровню инновационного развития экономики является ключевым фактором локализации кумулятивных форм организации производительных сил в субфедеральных экономических системах, сглаживание градиентов между которыми возможно путем идентификации и «выращивания» полюсов экономического роста и их последующей межтерриториальной экспансии.

Гипотеза 3. Повышения уровня конкурентоспособности и ускорения социально-экономического развития региона основано на использования инструментов и методов маркетингового позиционирования.

Гипотеза 1: Эффективное расположение элементов навигации позволит повысить скорость поиска розничного объекта, уменьшив временные затраты клиента, что отразится на возможности выбора местоположения торговой точки в малопривлекательных городских локациях с меньшей стоимостью арендной ставки.



Как будет выстроено научное исследование в связи с заявленной гипотезой



Объем данных, обеспечивающих достоверность результатов исследования

$$n = \frac{\sigma^2 t^2}{\Delta^2}$$

$$n = \frac{\sigma^2 t^2 N}{\sigma^2 t^2 + \Delta^2 N}$$

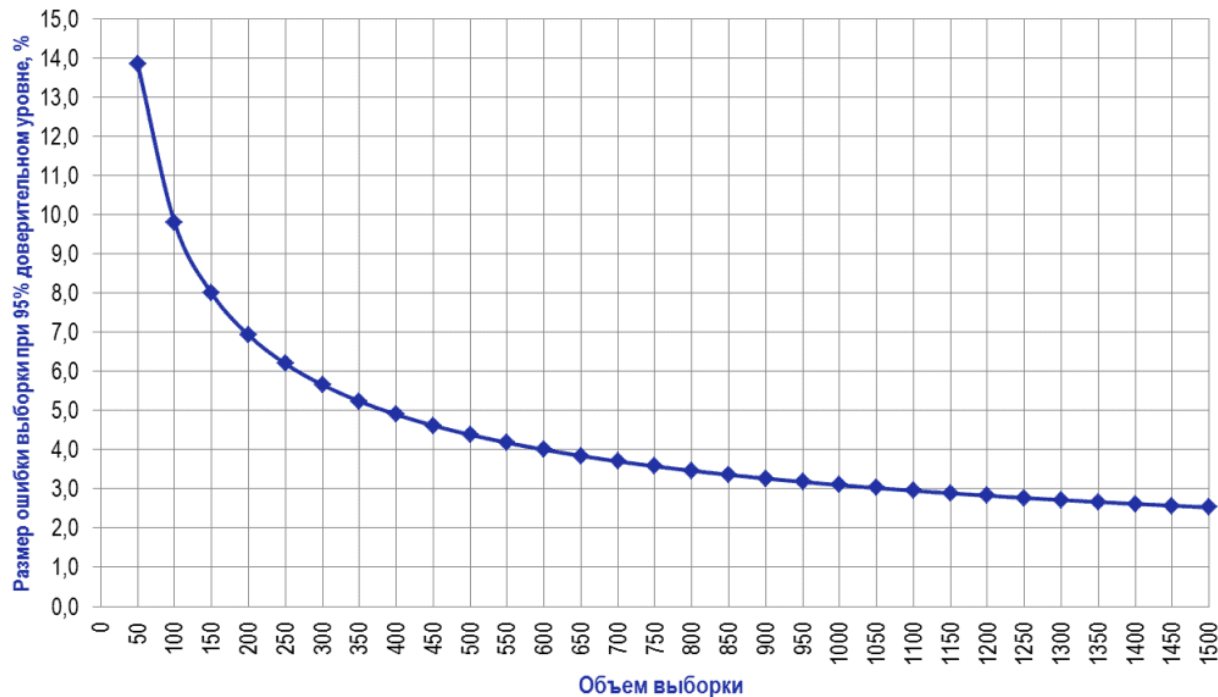
σ^2 – дисперсия изучаемого признака

t – коэффициент доверия согласно Р (доверительный уровень)

Δ - статистическая предельная ошибка выборки

N – объем генеральной совокупности

Зависимость ошибки выборки от ее объема



Калькулятор расчета n

КАЛЬКУЛЯТОРЫ ДЛЯ РАСЧЕТА СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ И ОПРЕДЕЛЕНИЯ
ЗНАЧИМОСТИ РАЗЛИЧИЙ:

КАЛЬКУЛЯТОР ДЛЯ РАСЧЕТА ДОСТАТОЧНОГО ОБЪЁМА ВЫБОРКИ

Доверительный уровень:

95% ▼

Ошибка выборки (?):

3 %

Объём генеральной совокупности (N):

(можно пропустить, если больше 100 000)

3000

Рассчитать

787

Пример определения оптимального объема выборки



Малые выборки ($n < 30$)

- Недостаточный объем данных для выявления статистически значимых различий или связей между переменными.
- Невозможность получить достоверных результатов исследования и выводов.
- В малых выборках больше вероятность случайных отклонений колебаний, которые искажают результаты и как результат неправильные выводы о связи.
- Малые выборки ограничивают использование множественной регрессии, факторного анализа, структурного моделирования.

Методы «малой выборки»

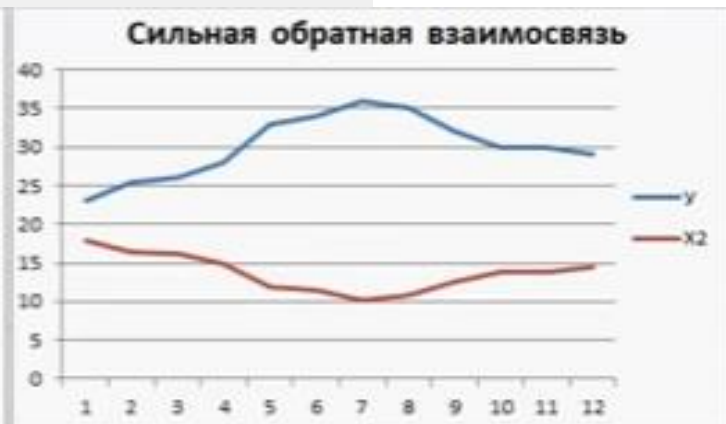
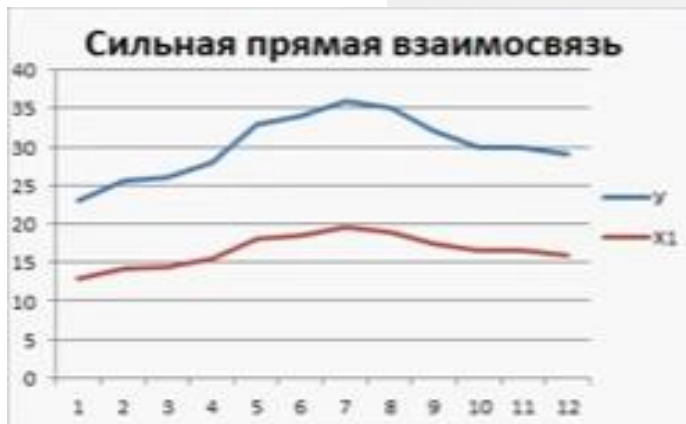
- критерий точной вероятности Фишера;
- двухфакторный непараметрический (ранговый) дисперсионный анализ Фридмана;
- коэффициент ранговой корреляции t Кендалла;
- коэффициент конкордации (W) Кендалла;
- H -критерий Краскела—Уоллеса для непараметрического (рангового) однофакторного дисперсионного анализа;
- U -критерий Манна—Уитни;
- медианный критерий;
- критерий знаков;
- коэффициент ранговой корреляции r Спирмена;
- t -критерий Уилкоксона.

Инструменты исследования и оценки связей

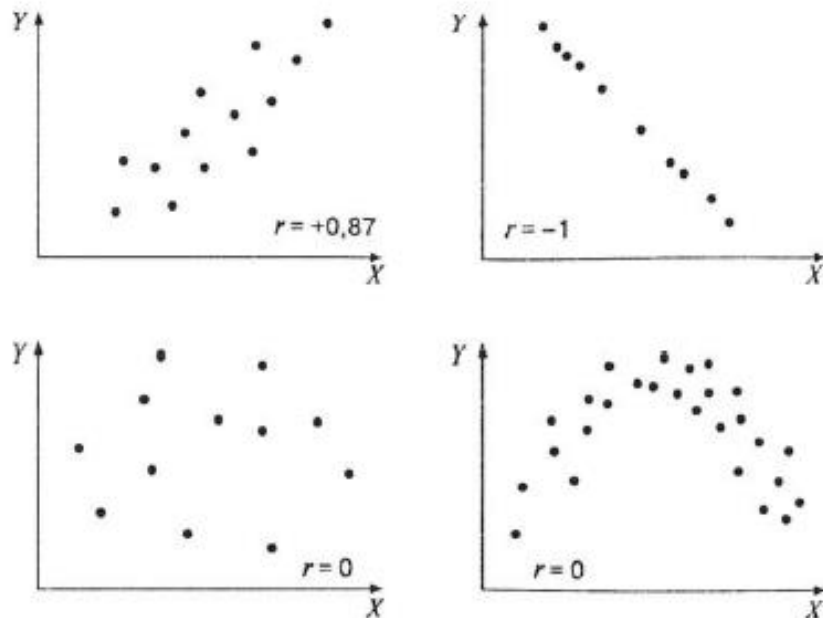
- Изучение связей между переменными, интересует исследователя с точки зрения отражения соответствующих **причинно-следственных отношений**.
- **Корреляционная зависимость** – это согласованные изменения двух (парная корреляционная связь) или большего количества признаков (множественная корреляционная связь). Суть ее заключается в том, что при изменении значения одной переменной происходит закономерное изменение (уменьшение или увеличение) другой(-их) переменной(-ых).
- **Корреляционный анализ** – статистический метод, позволяющий с использованием коэффициентов корреляции определить, существует ли зависимость между переменными и насколько она сильна.
- **Коэффициент корреляции** – двумерная описательная статистика, количественная мера взаимосвязи (совместной изменчивости) двух переменных.

Корреляционный анализ

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ



Примеры тесноты и направления связи



Примеры рассеивания и соответствующих коэффициентов корреляции

Критические значения коэффициента корреляции

n	p			
	0,1	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991
6	0,729	0,811	0,917	0,974
7	0,669	0,754	0,875	0,951
8	0,621	0,707	0,834	0,925
9	0,582	0,666	0,798	0,898
10	0,549	0,632	0,765	0,872
11	0,521	0,602	0,735	0,847
12	0,497	0,576	0,708	0,823
13	0,476	0,553	0,684	0,801
14	0,458	0,532	0,661	0,780
15	0,441	0,514	0,641	0,760
16	0,426	0,497	0,623	0,742
17	0,412	0,482	0,606	0,725
18	0,400	0,468	0,590	0,708
19	0,389	0,456	0,575	0,693

ДАННЫЕ ДЛЯ АНАЛИЗА

	ВРП на душу населения	Численность ЭАН	Инвестиции в основной капитал на душу населения	Поступление прямых иностранных инвестиций	Степень износа основных фондов	Число предприятий и организаций	Индекс промышленного производства	Уровень инновационной активности организаций
Белгородская область	506167,8	824	91978	846	47,6	38507	105,0	14,8
Брянская область	251085,7	613	45339	5	45,5	20515	104,1	6,2
Владимирская область	298375,9	733	56740	284	48,4	32376	103,1	9
Воронежская область	371924,5	1179	121500	172	40,4	57411	106,0	11,7
Ивановская область	181377,5	544	29752	85	48,3	31698	100,0	4,2
Калужская область	410481,4	539	87856	1544	41,9	26696	119,8	9
Костромская область	258533	327	33510	261	50,6	16464	107,0	2,8
Курская область	346101,2	573	90160	28	50,5	22347	103,2	5
Липецкая область	439070,1	599	123487	2226	55,8	22285	102,9	18,5
Московская область	506478,2	4097	93780	9423	45	239276	108,0	8,9
Орловская область	286468	380	58146	10	50,3	15195	96,4	6,8
Рязанская область	321083,3	541	55410	226	59,1	29801	97,5	12,1
Смоленская область	306413,8	523	60441	149	55,2	28220	103,6	6,5
Тамбовская область	288147,3	516	107116	11	60,6	17833	109,7	11
Тверская область	300328,7	688	78076	53	50,3	34273	102,8	8,7
Тульская область	372264,9	792	85959	833	47,6	35360	103,3	9,2
Ярославская область	403037,8	671	67640	452	54,9	45321	111,5	8,3
г. Москва	1260755	7195	161345	85326	37	1001076	100,7	14,3

Результаты корреляционного анализа

	ВРП на душу населения	Численность ЭАН	Инвестиции в основной капитал на душу населения	Поступление прямых иностранных инвестиций	Степень износа основных фондов	Число предприятий и организаций	Индекс промышленного производства	Уровень инновационной активности организаций
ВРП на душу населения	1,00							
Численность ЭАН	0,91	1,00						
Инвестиции в основной капитал на душу населения	0,75	0,63	1,00					
Поступление прямых иностранных инвестиций	0,95	0,92	0,62	1,00				
Степень износа основных фондов	-0,52	-0,57	-0,34	-0,52	1,00			
Число предприятий и организаций	0,95	0,96	0,62	0,99	-0,54	1,00		
Индекс промышленного производства	-0,03	-0,09	0,13	-0,16	-0,14	-0,14	1,00	
Уровень инновационной активности организаций	0,51	0,32	0,74	0,33	-0,01	0,32	-0,04	1,00

Расчет коэффициентов Пирсона и Чупрова

Успеваемость	Вид занятий				Итого
	МООК	МООК и рабочая	Занятия в аудитории	Занятия в аудитории и МООК	
меньше 3,5	10	7	2	3	22
3,5-3,9	12	13	27	12	64
4,0-4,5	12	15	24	60	111
более 4,5	15	17	20	80	132
Итого	49	52	73	155	329

Кп	0,395089119
Кч	0,248306317

Гипотеза: Введение цифровых технологий в образовательном пространстве повышает успеваемость студентов.

Не опровергается, так как $K_p > K_{\text{ч}}$, и оба коэффициента значительно отличаются от нуля. Кп более точно отображает результат исследования.

Школа молодого исследователя

Тема: «Алгоритм построения и проверки научной гипотезы»

Задать вопросы можно по почте:

Титова Ирина Николаевна,

titova@bsu.edu.ru

+ 7 910 360 00 93