

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАнных

**СЧЕТНАЯ ОБРАБОТКА
ПОЛУЧЕННЫХ ДАнных**

**ИЗМЕРЕНИЕ СВЯЗИ МЕЖДУ
ЯВЛЕНИЯМИ ИЛИ
ПРИЗНАКАМИ. КОРРЕЛЯЦИЯ**

Известно, что все явления в природе и обществе находятся во взаимной связи. При изучении связи между явлениями:

Во-первых, следует провести качественный анализ, позволяющий установить возможность связи между явлениями, определить ее характер, выяснить имеется ли между ними причинно-следственная связь, или оба из изучаемых явлений зависят от третьего.

Во-вторых, целесообразно статистически проверить наличие связи.

Для этих целей служит коэффициент корреляции, который одним числом дает оценку степени связи между изучаемыми явлениями, определяет ее силу и направление, позволяет распутать цепь причин и привлечь к их объяснению ряд различных факторов.

ОСНОВНЫЕ ВОПРОСЫ И КЛЮЧЕВЫЕ ПОНЯТИЯ

Различные явления или признаки могут иметь между собой две формы связи: **функциональную и корреляционную**.

Функциональная связь характерна для явлений, между которыми существует строгая зависимость, то есть изменение на определенную величину одного явления (признака) сопровождается строго определенным изменением другого явления (признака). При этом зависимость проявляется настолько сильно, что каждому определенному значению влияющего признака (x) всегда соответствует строго определенное, единственное значение результативного признака (y). Такая связь встречается только в идеальных условиях, когда предполагается, что никаких посторонних влияний нет.

Функциональная связь в основном имеет место там, где ее можно представить в виде уравнения, формулы (например, радиусу круга соответствует определенная площадь круга, степень расширения тела определяется температурой нагревания и т.п.).

В биологии и медицине, как правило, встречается корреляционная связь. **Корреляционная связь** – это связь, при которой от изменения одного явления зависит главным образом, но не исключительно, изменение другого явления, значения которого варьируют в определенных размерах вокруг своей средней величины. Например, известно, что масса тела человека зависит от изменения его роста.

Эти колебания массы тела варьируют в определенных размерах вокруг своей средней величины, т.к. связаны еще с питанием, физической нагрузкой, обменом веществ, состоянием здоровья и т.п.

Другими примерами могут служить, связь уменьшения уровня инфекционной заболеваемости с увеличением процента привитых лиц, взаимосвязь между цветом волос и цветом глаз (темные или светлые), зависимость заболеваемости раком кожи от степени пигментации кожных покровов, связь между уровнем вредных производных факторов и частотой профессиональных заболеваний.

По своему характеру корреляционная связь подразделяется на следующие виды:

- ***причинно-следственную;***
- ***обусловленную;***
- ***параллельную.***

Причинно-следственная связь проявляется в тех случаях, когда изменение одного явления (признака) является причиной изменения другого явления (признака). **Например,** увеличение числа диспансерных больных, получивших противорецидивное лечение, приводит к снижению частоты обострения хронических заболеваний, внедрение новых технологий на производстве приводит к изменению показателей производственного травматизма.

Обусловленная связь присутствует в тех случаях, когда изменение одного явления (признака) как бы создаст условия, способствующие изменению другого явления или признака. **Например,** высокая температура воздуха в летнее время может привести к росту частоты сердечно-сосудистых заболеваний, увеличению уровня пищевых отравлений среди населения, а понижение температуры воздуха в зимнее время сопровождается увеличением уровня заболеваемости населения простудными болезнями. Но при наличии ряда обстоятельств (например, проведение профилактических мероприятий) этого может не произойти.

Для параллельной связи характерно параллельное изменение двух явлений или признаков в зависимости от третьего. Например, изменение роста ребенка сопровождается изменением массы его тела и окружности груди. Но параметры таких изменений неодинаковы.

Различают прямолинейную и криволинейную корреляционную связь. Так, если при относительно равномерном изменении средних значений одного признака проходят равные изменения другого (например, соответствия между изменениями уровней максимального и минимального артериального давления) наблюдается прямолинейная корреляционная связь.

При криволинейной зависимости равномерное изменение одного признака сопровождается неравномерными, непропорциональными возрастающими и убывающими изменениями другого признака.

Методы корреляции можно применять только при измерении связи между различными признаками только лишь в качественно однородной совокупности. Нельзя, например, сопоставлять рост и массу тела лиц разного пола и возраста.

Способами изображения связи могут быть:

- **таблицы,**
- **рисунки (графики)**
- **коэффициенты корреляции.**

Изучаемые явления (признаки) могут быть выражены количественно или описательно, представлены сгруппированными или несгруппированными данными.

Выразить количественно – это значит выразить числом (например, рост – в сантиметрах или метрах, массу тела – в килограммах, содержание железа в крови – в мг%, частоту заболеваний в промиллях или продецимиллях), а выразить описательно – значит выразить словесно.

Описательные признаки – обычно альтернативные признаки. Альтернативным называют случай, когда вариация исчерпывается двумя возможностями: заболел или не заболел, привит или не привит, выздоровел или не выздоровел.

Сгруппированные данные – это данные, объединенные в группы по их величине в пределах одного интервала. В виде сгруппированных данных могут быть представлены значения одного из изучаемых явлений (признаков) или значения обоих признаков: влияющего (x) и результативного (y). Чаще всего группируют данные при большом числе наблюдений.

Несгруппированные данные – это данные каждого из изучаемых признаков, выраженные одним числом.

В **таблицах** коррелируемые данные представляют в виде *корреляционных рядов или корреляционной решетки*.

Корреляционный ряд – это ряд числовых изменений определенного явления (признака). В корреляционной таблице минимально может быть два таких ряда. Один из них (x) относится к влияющему признаку, а другой (y) – к результативному.

Схема корреляционного ряда

V_x	V_y
X_1	Y_1
X_2	Y_2
\vdots	\vdots
X_n	Y_n

Корреляционная решетка – таблица, в которой каждому значению влияющего признака (x) относится несколько значений результативного признака (y). При этом каждому сочетанию x и y соответствует определенное поле .

Схема корреляционной решетки

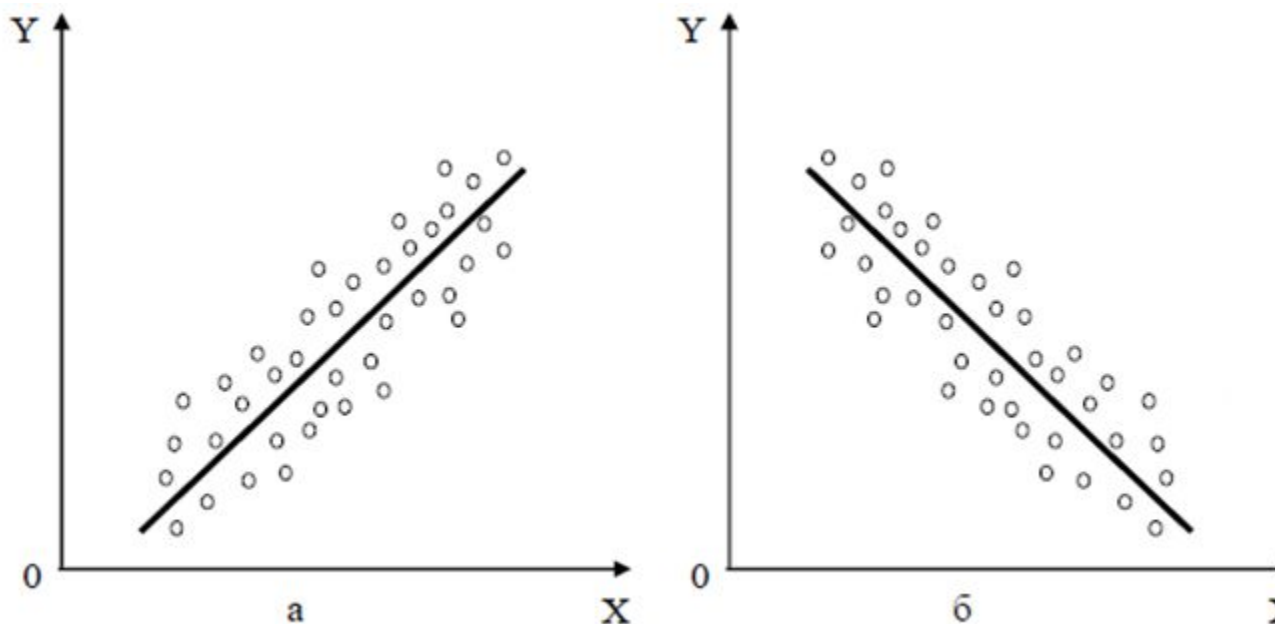
V_x	V_y			
	Y_1	Y_2	...	Y_n
X_1				
X_2				
⋮				
X_n				

Примечание:

X_1, X_2, X_3 и т.д. – разновидности одного явления (признака);

Y_1, Y_2, Y_3 и т. д. – разновидности другого явления (признака).

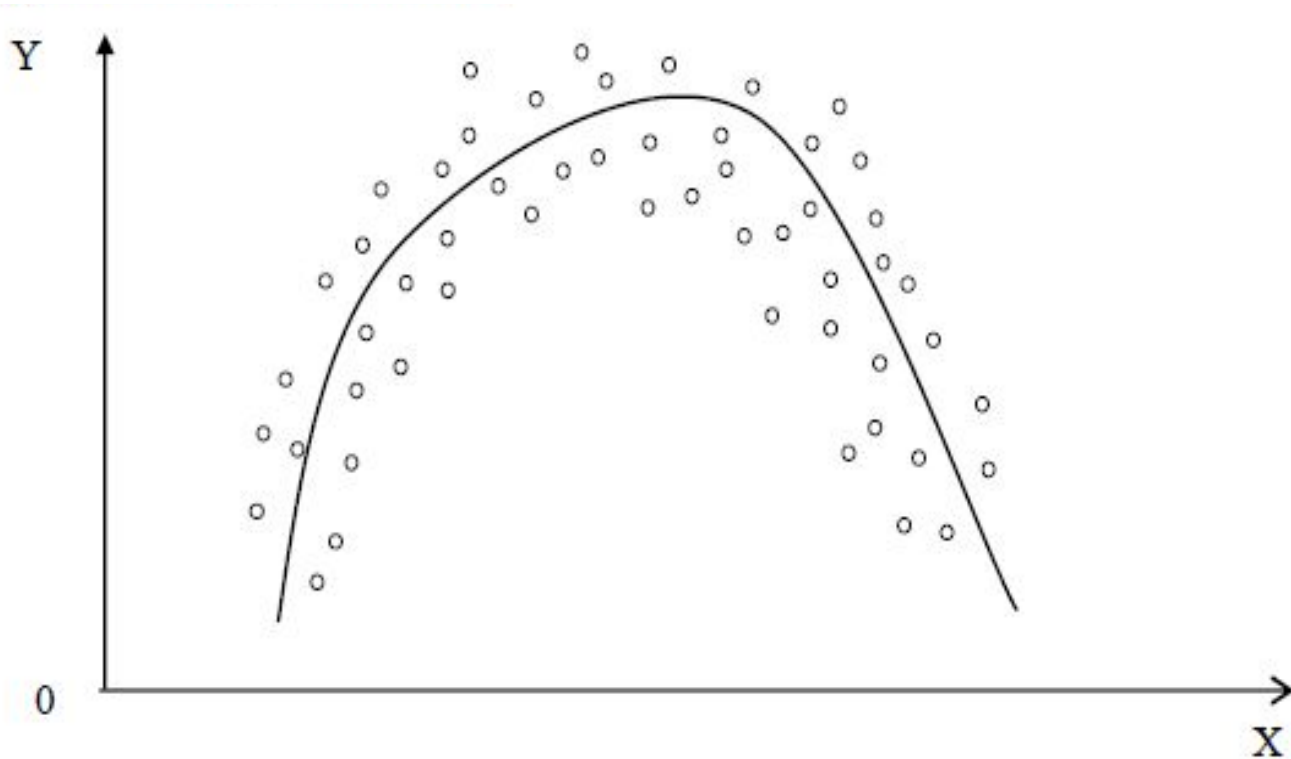
Графики (рисунки) дают представление о наличии корреляционной связи и ее направлении.



Линейная корреляция между величинами X и Y

а – положительная, прямая; б – отрицательная, обратная.

Распределение точек на рисунке показывает, что они широко рассеяны вокруг линии. Их общее направление соответствует прямой линии. Тенденции к образованию кривой незаметно. Средние величины весьма близко подходят к прямой линии. В этом случае мерилom связи должен быть выбран коэффициент корреляции.



Криволинейная зависимость между величинами X и Y

Если размещение точек на графике грубо похоже, по очертанию на перевернутую букву U и прямой линией описать его было бы невозможно, тогда коэффициент корреляции применять не следует. Для этого существуют другие методы анализа.

Таблицы и графики дают лишь представление о наличии и направлении связи. Однако измерить и оценить статистическую достоверность этой связи можно при помощи **коэффициента корреляции**.

Существует много способов вычисления таких коэффициентов. Выбор одного какого-либо из них находится в зависимости от следующих факторов:

- 1. Каким образом выражены признаки изучаемых явлений, находящихся во взаимной связи, – описательно или количественно.
- 2. В каком виде представлены данные – сгруппированном или несгруппированном.
- 3. Требуется ли измерить связь между двумя явлениями или между несколькими.

Наиболее простыми для расчета и широко используемыми являются:

- коэффициент линейной корреляции (r_{xy});
- коэффициент ранговой корреляции (ρ);
- коэффициент ассоциации (Q).

По направлению связь может быть прямой и обратной.

Прямой (положительной) связью называется такая связь, когда оба явления изменяются в одном направлении. Например, с ухудшением питания ухудшаются показатели здоровья населения. Обозначается знаком плюс (+)

Обратной (отрицательной) связью называется связь, когда одно явление увеличивается, а другое при этом уменьшается. Обозначается знаком минус (-).

По силе связи коэффициенты корреляции колеблются от единицы (полная функциональная связь) до нуля (отсутствие связи). Промежуточные значения ($1 > r > 0$) говорят о наличие связи более или менее сильной степени. Чем больше среднему значению одного признака соответствует значений другого признака, тем выше сила связи между ними. Оценка размеров корреляции может производиться по следующей схеме.

Сила связи	Величина коэффициента при наличии	
	прямой корреляции	обратной корреляции
Малая (низкая, слабая)	от 0 до +0,3	от 0 до -0,3
Средняя	от 0,3 до +0,7	от -0,3 до -0,7
Большая (высокая, сильная)	от 0,7 до +1,0	от -0,7 до -1,0

1. КОЭФФИЦИЕНТ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ

Наиболее точный. Его используют:

- при прямолинейной зависимости между изучаемыми явлениями (признаками);
- небольшом числе наблюдении (число парных вариант не более 30);
- несгруппированных данных.

Коэффициент корреляции (r_{xy}) одним числом измеряет силу связи между изучаемыми явлениями и дает представление о ее направлении;

2. КОЭФФИЦИЕНТ РАНГОВОЙ КОРРЕЛЯЦИИ

Применяют в том случае, когда не нужна большая точность оценки силы связи.

Его используют:

- при прямолинейной, криволинейной зависимости между изучаемыми явлениями или когда определить ее вид невозможно;
- небольшом числе наблюдений (не больше 30);
- при сравнении не только количественных, но и качественных (атрибутивных) признаков;
- при измерении связи между альтернативными признаками, имеющими несколько градаций;
- когда в сравниваемых рядах величин имеются открытые группы, не имеющие четких границ (например, в ряду величин, характеризующих стаж работы, могут быть группы: «до 5 лет», «более 20 лет»).

3. КОЭФФИЦИЕНТ АССОЦИАЦИИ

Является наиболее точным и находит применение для измерения связи между альтернативными признаками.

Названные коэффициенты рассчитываются по следующим формулам:

Линейной корреляции	Ранговой корреляции	Ассоциации
$r_{xy} = \frac{\sum d_x \times d_y}{\sqrt{\sum d_x^2 \times d_y^2}}$ <p>где: r_{xy} – коэффициент корреляции X и Y – числовые значения сравниваемых явлений признаков d_x и d_y – отклонения каждого значения ряда «x» или ряда «y» от средней величины данного ряда \sum – знак суммы</p>	$\rho_{xy} = 1 - \frac{\sqrt{6 \sum d^2}}{n \times (n^2 - 1)}$ <p>где: ρ_{xy} – коэффициент корреляции рангов n – число парных членов коррелируемых рядов (x и y) d – разность между порядковыми номерами парных членов рядов \sum – знак суммы</p>	$Q = \frac{ad - bc}{ad + bc}$ <p>где: Q – коэффициент ассоциации a, b, c, d – обозначения каждого из 4-х сравниваемых показателей</p>

Полученные коэффициенты нуждаются в подтверждении их достоверности. Для этого необходимо вычислить среднюю ошибку коэффициента корреляции.

Средняя ошибка (m) коэффициентов линейной и ранговой корреляции вычисляется по однотипной формуле:

для коэффициента линейной корреляции	для коэффициента ранговой корреляции
$m_r = \pm \sqrt{\frac{1-r_{xy}^2}{n-2}}$ <p>где: m_r – средняя ошибка n – число пар r_{xy}^2 – коэффициент линейной корреляции</p>	$m_\rho = \pm \sqrt{\frac{1-r_{xy}^2}{n-2}}$ <p>где: m_ρ – средняя ошибка n – число пар r_{xy}^2 – коэффициент ранговой корреляции</p>

Средняя ошибка коэффициента ассоциации определяется по формуле:

$$m_Q = \frac{1-Q}{2} \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}},$$

Достоверность коэффициентов корреляционной связи определяется с помощью критерия достоверности – t, который вычисляется путем деления величины показателя на величину рассчитанной для него средней ошибки:

$$t_r = \frac{r_{xy}}{m_r} \quad \text{или} \quad t_\rho = \frac{\rho}{m_\rho} \quad \text{или} \quad t_Q = \frac{Q}{m_Q}$$

Коэффициенты корреляционной связи считаются **достоверными**, если они в 3 (три) раза превышают свою среднюю ошибку;

При малом числе наблюдений значение критерия достоверности, рассчитанное для конкретного коэффициента, сравнивается с критериями в специальной таблице (**по Н.А. Плохинскому**), соответствующими числу наблюдений в данном исследовании.

Значение критерия tr для трех степеней вероятности (по Н.А. Плохинскому)

p n	95%	99%	99,9%	p n	95%	99%	99,9%
1	12,7	63,7	37,0	10	2,2	3,2	4,6
2	4,3	9,9	31,6	11	2,2	3,1	4,4
3	3,2	5,8	12,9	12	2,2	3,1	4,3
4	2,8	4,6	8,6	13	2,2	3,0	4,1
5	2,6	4,0	6,9	14-15	2,1	3,0	4,1
6	2,4	3,7	6,0	16-17	2,1	2,9	4,0
7	2,4	3,5	5,3	18-20	2,1	2,9	3,9
8	2,3	3,4	5,0	21-24	2,1	2,8	3,8
9	2,3	3,3	4,8	25-29	2,0	2,8	3,7

Оценка достоверности коэффициента корреляции может осуществляться по специальной таблице (при малых выборках) **без предварительных расчетов m и t** . Необходимо лишь сравнить r_{xy} со стандартным коэффициентом корреляции, рассчитанным и представленным в таблице для различной степени вероятности и различного числа наблюдений.

Стандартные коэффициенты корреляции, которые считаются

Число степеней свободы, $n' (n - 2)$	Уровень вероятности P (в %)			Число степеней свободы, $n' (n - 2)$	Уровень вероятности P (в %)		
	95,0	98,0	99,0		95,0	98,0	99,0
1	0,997	0,999	0,999	12	0,532	0,612	0,661
2	0,950	0,980	0,990	13	0,514	0,592	0,641
3	0,878	0,934	0,959	14	0,497	0,574	0,623
4	0,811	0,882	0,917	15	0,482	0,558	0,606
5	0,754	0,833	0,874	16	0,468	0,542	0,590
6	0,707	0,789	0,834	17	0,456	0,528	0,575
7	0,666	0,750	0,798	18	0,444	0,516	0,561
8	0,632	0,716	0,765	19	0,433	0,503	0,549
9	0,602	0,685	0,735	20	0,423	0,492	0,537
10	0,576	0,658	0,708	25	0,381	0,445	0,487
11	0,553	0,634	0,684	30	0,349	0,409	0,449

МЕТОДИКА ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА ЛИНЕЙНОЙ КОРРЕЛЯЦИИ

1-й способ Алгоритмы расчета коэффициента линейной корреляции (r_{xy}), его ошибки (m_r) и коэффициента достоверности (t_r)

V_x	V_y	d_x <small>(X-M_x)</small>	d_y <small>(Y-M_y)</small>	$d_x \times d_y$	d_x^2	d_y^2	Формулы
							$r_{xy} = \frac{\sum d_x \times d_y}{\sqrt{\sum d_x^2 \times d_y^2}}$
$\sum V_x$	$\sum V_y$						$m_r = \pm \sqrt{\frac{1 - r_{xy}^2}{n - 2}}$
$M_x = \frac{\sum V_x}{n}$	$M_y = \frac{\sum V_y}{n}$			$\sum d_x \times d_y$	$\frac{\sum d_x^2}{2}$	$\frac{\sum d_y^2}{2}$	$t_r = \frac{r_{xy}}{m_r}$

1. Записать исходные данные в виде двух вариационных рядов (графы 1 и 2)
2. Найти суммы вариантов в каждом вариационном ряду (Σx и Σy) и определить средние арифметические величины (Mx и My) – графы 1 и 2.
3. Найти dx и dy – отклонения каждой варианты от средних величин (графы 3 и 4).
4. Полученные отклонения перемножить попарно ($dx \times dy$) и найти сумму полученных произведений ($\Sigma dx \times dy$) – графа 5.
5. Каждое отклонение в обоих рядах возвести в квадрат и определить сумму квадратов отклонений ряда Vx (графа 6) и ряда Vy (графа 7).
6. Определить произведение $\Sigma d^2x \times \Sigma d^2y$ и из произведения извлечь квадратный корень $\sqrt{\Sigma d^2x \times \Sigma d^2y}$
7. Подставить полученные данные в формулу и рассчитать коэффициент корреляции (R_{xy}) – графа 8.
8. Подставить необходимые данные в формулу и рассчитать среднюю ошибку (mr) коэффициента корреляции – графа 8.
9. Подставить необходимые данные в формулу и рассчитать коэффициент достоверности (tr) – графа 8.

2-й способ

При наличии вычислительной техники расчет коэффициента линейной корреляции производится по следующей схеме

Алгоритм расчета (на ЭВМ) коэффициента корреляции (R_{xy})

V_x	V_y	$V_x \cdot V_y$	V_x^2	V_y^2	Формула
					$r_{xy} = \frac{n \times \sum V_x V_y - \sum V_x \times \sum V_y}{\sqrt{[n \times \sum V_x^2 - (\sum V_x)^2] \times [n \times \sum V_y^2 - (\sum V_y)^2]}}$
$\sum V_x$	$\sum V_y$	$\sum V_x \times V_y$	$\sum V_x^2$	$\sum V_y^2$	

ПРИМЕР. Определение r_{xy} между температурой тела и частотой пульса в минуту

1-й способ.

Температура тела($^{\circ}\text{C}$), x	Частота пульса (удары в минуту), y	d_x ($X-M_x$)	d_y ($Y-M_y$)	$d_x \times d_y$	d_x^2	d_y^2
36	60	-2	-20	40	4	400
36	70	-2	-10	20	4	100
38	80	0	0	0	0	0
40	90	+2	+10	20	4	100
40	100	+2	+20	40	4	400
$\sum x = 190$ $M_x = \frac{190}{5} = 38$	$\sum y = 400$ $M_y = \frac{400}{5} = 80$					

Рассчитаем коэффициент корреляции:

$$r_{xy} = \frac{\sum d_x \times d_y}{\sqrt{\sum d_x^2 \times d_y^2}} = \frac{120}{\sqrt{16 \times 1000}} = \frac{120}{126.49} = +0,949$$

Рассчитаем среднюю ошибку (m_r) коэффициента достоверности:

$$m_r = \pm \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \pm \sqrt{\frac{1 - 0,949^2}{5 - 2}} = \pm \sqrt{\frac{1 - 0,901}{3}} = \pm \sqrt{\frac{0,099}{3}} = \pm 0,18$$

Рассчитаем коэффициент достоверности

$$t_r = \frac{r_{xy}}{m_r} = \frac{0,949}{0,18} = 5,27$$

Оценим критерий t_r по специальным таблицам (**по Н.А. Плохинскому**) и (**по Л.С. Каменскому**)

Вывод. Коэффициент корреляции, равный 0,949, достоверен с вероятностью безошибочного прогноза $p > 95\%$, так как при $n = 3$ ($5 - 2$) полученный нами критерий t будет больше $t_{табл.} = 3,2$ ($p = 95\%$) и меньше $t_{табл.} = 5,8$ ($p = 99\%$). Материалы выборочного исследования позволяют утверждать, что в генеральной совокупности существует сильная прямая связь между температурой тела и частотой пульса.

**2-й
способ**

№ п/п	V_x	V_y	$V_x \times V_y$	V_x^2	V_y^2	Расчет по формуле
1	36	60	2160	1296	3600	$r_{xy} = \frac{5 \times 15320 - 190 \times 400}{\sqrt{(5 \times 7236 - 190^2) \times (5 \times 3300 - 400^2)}} = \frac{600}{\sqrt{400000}} =$ $= \frac{600}{632,46} = +0,949$
2	36	70	2520	1296	4900	
3	38	80	3040	1444	6400	
4	40	90	3600	1600	8100	
5	40	100	4000	1600	10000	
	190	400	15320	7236	33000	

Вывод. Между температурой тела и частотой пульса имеется прямая и сильная связь.

МЕТОДИКА КОЭФФИЦИЕНТА РАНГОВОЙ КОРРЕЛЯЦИИ

Алгоритм расчета коэффициента ранговой корреляции (ρ), его ошибки ($m\rho$) и коэффициента достоверности ($t\rho$)

Признак x	Признак y	Порядковые номера (ранги) для рядов		Разность между рангами ($d=x'-y'$)	Квадрат разности (d^2)	Формулы для расчета
		x'	y'			
						$\rho = 1 - \frac{6 \times \sum d^2}{n \times (n^2 - 1)}$ $m\rho = \pm \sqrt{\frac{1 - \rho^2}{n - 2}}$ $t\rho = \frac{\rho}{m\rho}$
				$\sum d^2$		

1. Составить ряды из парных признаков (графы 1 и 2).
2. Каждую величину признака заменить ранговым (порядковым) номером – x' и y' . (в тех случаях, когда имеется несколько одинаковых по величине чисел, порядковый номер обозначают средним числом из суммы очередных порядковых их номеров). Ранжировать значения обоих рядов в строго определенном направлении от меньшей величины к большей или от большей к меньшей (графы 3 и 4).
3. Определить разность между рангами для каждой пары членов ряда (по каждой строке) – графа 5.
4. Возвести в квадрат каждое из полученных значений разности между рангами и определить сумму квадратов разности рангов ($\sum d^2$) – графа 6.
5. Подставить полученные данные в формулу и рассчитать коэффициент корреляции рангов – графа 7.
6. Подставить необходимые данные в формулу и рассчитать среднюю ошибку (mp) коэффициента ранговой корреляции – графа 7.
7. Подставить необходимые данные в формулу и рассчитать коэффициент достоверности (tp) – графа 7.

ПРИМЕ

Данные о заболеваемости дифтерией жителей городов Н-ской области и о выполнении плана профилактических прививок в отчетном году

Города	Заболеваемость дифтерией (на 100 чел.)	Выполнение плана профпрививок (в %)	Ранги		Разность рангов	Квадрат разности рангов
			X_i	Y_i		
	x	y			$d (x-y)$	d^2
А	3,29	86	1	8	7	49
Б	3,21	82	2	10	8	64
В	3,17	88	3	7	4	16
Г	2,85	84	4	9	5	25
Д	2,84	90	5	6	1	1
Е	2,71	95	6	4	2	4
Ж	2,23	100	7	1	6	36
З	1,82	92	8	5	3	9
И	1,80	98	9	2	7	49
К	0,70	96	10	3	7	49
						$\sum d^2=302$

Рассчитываем коэффициент ранговой корреляции:

$$\rho = 1 - \frac{6 \times \sum d^2}{n \times (n^2 - 2)} = 1 - \frac{6 \times 302}{10 \times (10^2 - 1)} = 1 - \frac{1812}{990} = 1 - 1,83 = -0,83$$

Рассчитаем среднюю ошибку

(

$$m_{\rho} = \pm \sqrt{\frac{1 - \rho_{xy}^2}{n - 2}} = \pm \sqrt{\frac{1 - (-0,83^2)}{10 - 2}} = \pm \sqrt{\frac{1 - 0,689}{8}} = \pm \sqrt{0,0389} = \pm 0,197$$

Рассчитаем коэффициент
достоверности

$$t_{\rho} = \frac{\rho}{m_{\rho}} = \frac{-0,83}{0,197} = 4,2$$

Условие достоверности коэффициента корреляции рангов

$$t_{\rho} \geq 3 \times m_{\rho}$$

Таким образом, $-0,83 > 3 \times 0,197$ или $-0,83 > 0,591$ т.е. полученные результаты достоверны.

Вывод. Полученный коэффициент корреляции рангов статистически достоверен. Следовательно, можно говорить о сильной, обратной связи между заболеваемостью дифтерией и процентом выполнения плана профилактических прививок.

Заболеваемость выше в тех городах, где план профилактических прививок выполнен недостаточно.

МЕТОДИКА ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА АССОЦИАЦИИ

Алгоритм расчета коэффициента ассоциации (Q), его ошибки (ma) и коэффициента достоверности (ta).

X \ Y	Y_1 Заболело	Y_2 Не заболело
X_1 Вакцинировано	a	b
X_2 Невакцинировано	c	d

1. Построить четырехпольную таблицу. В первом столбце этой таблицы наносят обе разновидности одного явления – X_1 и X_2 , а в первой строке – обе разновидности второго – Y_1 и Y_2 . При этом X_1 и Y_1 обозначают положительные разновидности (например, выздоровевшие, иммунизированные и т. д.), а X_2 и Y_2 – отрицательные (например, не выздоровевшие, не иммунизированные). Обозначить через буквы a , b , c , d четыре поля, в которые внести исходные данные:

2. Вычислить произведения ad и bc .

3. Подставить полученные данные в формулу: $Q = (ad - bc)/(ad + bc)$ и рассчитать коэффициент ассоциации.

4. Подставить необходимые данные в формулу:

$$m_Q = \frac{1-Q}{2} \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

и рассчитать среднюю ошибку коэффициента ассоциации.

5. Подставить необходимые данные в формулу:

$$t_Q = \frac{Q}{m_Q}$$

и рассчитать коэффициент достоверности.

ПРИМЕ

Необходимо определить, влияет ли вакцинация против гриппа на заболеваемость от этой инфекции. Известно, что на промышленном предприятии с общей численностью в 2000 человек 1200 рабочим были сделаны прививки против гриппа, а 800 рабочих остались не привитыми. Заболело из привитых 240 человек, а из непривитых – 320.

Распределение заболевших и не заболевших среди привитых на промышленном предприятии

	Заболело	Не заболело	Всего
Привитых	240	960	1200
Не привитых	320	480	800
Всего	560	1440	2000

Вычислим произведения ad и bc :

$$a \times d = 240 \times 480 = 11520$$

$$b \times c = 960 \times 320 = 307200$$

Подставим полученные данные в формулу и рассчитаем коэффициент ассоциации:

$$Q = \frac{ad - bc}{ad + bc} = \frac{11520 - 307200}{11520 + 307200} = \frac{-192000}{422400} = -0,45$$

Рассчитаем среднюю ошибку (m):

$$m_Q = \frac{1 - Q}{2} \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \frac{1 - (-0,45)}{2} \times \sqrt{\frac{1}{240} + \frac{1}{960} + \frac{1}{320} + \frac{1}{480}} = 0,725 \times \sqrt{0,0104} = 0,725 \times 0,102 = 0,074$$

Рассчитаем коэффициент достоверности:

$$t_Q = \frac{Q}{m_Q} = \frac{-0,45}{0,074} = -6,1$$

Таким образом, $-0,45 > 3 \times 0,074 > 0,222$. Коэффициент статистически достоверен, что означает, что существует обратная, средней силы связи между заболеваемостью гриппом и вакцинацией против него.