

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Северо-Осетинская государственная медицинская академия»
Министерства здравоохранения Российской Федерации**

КАФЕДРА ОБЩЕСТВЕННОГО ЗДОРОВЬЯ И ЗДРАВООХРАНЕНИЯ

Туаева И.Б., Бадоева З.А., Габараева Л.Н., Гудцова А.П.

КОРРЕЛЯЦИЯ. СТАНДАРТИЗАЦИЯ

Учебно-методическое пособие по медицинской статистике

Владикавказ, 2018

СОСТАВИТЕЛИ:

Туаева И.Б. – к.м.н., доцент, зав. кафедрой общественного здоровья и здравоохранения ФГБОУ ВО СОГМА Минздрава России

Бадоева З.А. – к.м.н., доцент кафедры общественного здоровья и здравоохранения ФГБОУ ВО СОГМА Минздрава России

Габараева Л.Н. – к.м.н., старший преподаватель кафедры общественного здоровья и здравоохранения ФГБОУ ВО СОГМА Минздрава России

Гудцова А.П. – к.м.н., доцент кафедры общественного здоровья и здравоохранения ФГБОУ ВО СОГМА Минздрава России

РЕЦЕНЗЕНТЫ:

А.Р. Аликова – д.м.н., профессор, зав. кафедрой гуманитарных, социальных и экономических наук ФГБОУ ВО СОГМА Минздрава России

Р.В. Золоев – д.м.н., профессор, зав. кафедрой ортопедической стоматологии, пропедевтики и постдипломного образования ФГБОУ ВО «Северо-Осетинский государственный университет им. К.Л. Хетагурова»

Учебно-методическое пособие предназначается для подготовки к практическим занятиям, аудиторной и внеаудиторной самостоятельной работе обучающихся. В пособии представлены методы медицинской статистики, широко применяемые как при изучении показателей здоровья, так и при анализе деятельности медицинских учреждений, системы охраны, укрепления и восстановления здоровья населения.

Пособие направлено на формирование у обучающихся базовых компетенций обработки и анализа информации, выраженной числовыми данными.

Учебный материал охватывает основные вопросы корреляции и стандартизации формирует у обучающихся системное представление о возможностях и особенностях применения статистических методов для выявления закономерностей развития различных общественных явлений. Даны типовые задачи, образцы их решения, задачи для самостоятельной работы, тесты.

Учебно-методическое пособие составлено в соответствии с учебной программой дисциплины.

Утверждено и рекомендовано к печати Центральным координационным учебно-методическим советом ФГБОУ ВО СОГМА Минздрава России

Протокол № 4 от 16.05.2018 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
Глава 1. КОРРЕЛЯЦИЯ.....	5
1.1. Корреляция, понятие, формы связи	5
1.2. Коэффициент корреляции	5
1.3. Методы вычисления коэффициента корреляции	6
1.4. Коэффициент ранговой корреляции Спирмена	6
1.5. Метод квадратов (метод Пирсона)	7
1.6. Вычисление ошибки коэффициента корреляции	8
ТИПОВЫЕ ЗАДАЧИ	9
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	16
ПРИЛОЖЕНИЯ.....	17
Глава 2. СТАНДАРТИЗАЦИЯ.....	19
2.1. Стандартизация, понятие. Методы стандартизации	19
2.2. Прямой метод стандартизации	19
2.3. Этапы расчета стандартизованных показателей	20
ТИПОВЫЕ ЗАДАЧИ	21
ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	29
ЛИТЕРАТУРА (основная, дополнительная)	31

ВВЕДЕНИЕ

Современный этап развития общества характеризуется широким внедрением статистики в различные области науки, техники, народного хозяйства. Трудно назвать область, где статистика не могла бы найти себе применения. Это в полной мере относится к медицине и здравоохранению.

С помощью количественных характеристик, с учетом конкретных исторических условий статистика помогает выявить важнейшие закономерности различных процессов в экономической, социальной жизни общества, в его здоровье, а также в системе организации медицинской помощи населению.

Одним из важнейших элементов исследований является проведение научного анализа полученных данных на основе использования статистических методов.

Деятельность врачей разных специальностей неизменно связана с учетом, разработкой и анализом статистических материалов. Умение обобщить, проанализировать полученную в повседневной медицинской практике информацию позволяет на высшем качественном уровне подходить к решению клинических и организационных проблем. Кроме того, нередко врачу приходится самому проводить научные статистические исследования, потому изучение статистического метода при подготовке врачей имеет большое значение в системе высшего медицинского образования.

Изучение статистических методов способствует развитию у врачей критических взглядов, дедуктивных и индуктивных способностей (дедукция – метод анализа, при котором из общих положений логически выводятся частные; индукция – метод анализа от частного к общему). Статистический анализ позволяет обосновать ту или иную тактику врача в предупреждении и лечении заболеваний.

Настоящее пособие предназначается для практических занятий, внеаудиторной и самостоятельной работы и включает: введение, цели и задачи изучения темы, задания для самостоятельной работы обучающихся, контрольные вопросы, тестовые задания, ситуационные задачи и список обязательной и рекомендуемой литературы.

Глава 1 КОРРЕЛЯЦИЯ.

1.1. Корреляция, понятие, формы связи.

При изучении общественного здоровья и здравоохранения в научных и практических целях исследователю часто приходится проводить статистический анализ связей между факторными и результативными признаками статистической совокупности (причинно-следственная связь) или определение зависимости параллельных изменений нескольких признаков этой совокупности от какой-либо третьей величины (от общей их причины). Необходимо уметь изучать особенности этой связи, определять ее размеры и направление, а также оценивать ее достоверность. Для этого используются методы корреляции.

Корреляция (лат. слово) – это соотношение, взаимосвязь между признаками.

Различают 2 формы связи: функциональную и корреляционную.

Функциональная связь характеризуется тем, что каждому значению одного признака соответствует строго определенное значение другого признака и изменение величины одного признака вызывает совершенно определенные изменения величины другого признака, т.е. функциональная связь – отражает строгую зависимость процессов или явлений, изменение какого-либо одного явления обязательно связано с изменением другого явления на определенную величину (объем газа и давление, площадь круга зависит от радиуса круга и т.д.). Эта связь характерна для физико-химических процессов и присуща неживой природе.

Корреляционная связь – это связь между явлениями, проявляющаяся не в каждом отдельном случае, а при массовом сопоставлении рассматриваемых признаков.

Итак, **корреляционная связь** — это такая связь, при которой каждому определенному значению одного признака соответствует несколько значений другого взаимосвязанного с ним признака (связь между ростом и массой тела человека; связь между температурой тела и частотой пульса и др.), поэтому она проявляется лишь при массовом сопоставлении признаков в количественно однородной совокупности и характерна для социально-гигиенических и медико-биологических процессов.

Признаки могут быть качественными и количественными несгруппированными величинами (абсолютными и производными). Главным является установление причинных взаимосвязей, подтверждающих зависимость одного явления от другого или от какой-то общей причины. С этой целью определяют коэффициент корреляции, который позволяет оценить характер, силу и достоверность взаимосвязи изучаемых признаков.

1.2. Коэффициент корреляции.

Коэффициент корреляции – это величина, характеризующая направление и силу связи между признаками, который одним числом дает представление о направлении и силе связи между признаками (явлениями), пределы его колебаний от 0 до ± 1 .

По направлению связь между явлениями может быть прямая (+) и обратная (-).

Прямая связь (положительный коэффициент корреляции) – с увеличением одного признака увеличивается другой признак (+). Например, чем старше ребенок, тем больше его рост; по мере снижения температуры тела, как правило, частота пульса уменьшается и т.д.

Обратная связь (отрицательный коэффициент корреляции) – с увеличением одного признака (явления) другой уменьшается (-).

Под *теснотой* (силой) *связи* понимают степень сопряженности между признаками. Чем больше среднему значению одного признака соответствует среднее значение другого, тем больше теснота, сила связи между ними. Теснота связи определяется величиной коэффициента корреляции от 0 до ± 1 (табл. 1).

В зависимости от численного выражения коэффициента корреляции различают связь слабую (0,0 до 0,3), среднюю (от 0,3 до 0,7) и сильную от 0,7 до 1,0 (см. табл. 1).

Таблица 1.

Определение тесноты и направления связи по коэффициенту корреляции.

ОЦЕНКА КОРРЕЛЯЦИИ	ВЕЛИЧИНА КОЭФФИЦИЕНТА ПРИ НАЛИЧИИ	
	Прямой корреляции (+)	Обратной корреляции (-)
Связь отсутствует	0	0
Малая (низкая, слабая)	От 0 до +0,3	От -0,3 до 0
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
	+1,0	-1,0

Корреляционная связь может быть прямолинейной и криволинейной.

Прямолинейная связь – характеризуется относительно равномерным изменением средних значений одного признака при равных изменениях другого.

Криволинейная связь – при равномерном изменении одного признака могут наблюдаться возрастающие или убывающие значения другого признака.

Практическое значение установления корреляционной связи.

Выявление причинно-следственной связи между факторными и результативными признаками (при оценке физического развития, для определения связи между условиями труда, быта и состоянием здоровья, при определении зависимости частоты случаев болезни от возраста, стажа, наличия производственных вредностей и др.).

Зависимость параллельных изменений нескольких признаков от какой-то третьей величины. Например, под воздействием высокой температуры в цехе происходят изменения кровяного давления, вязкости крови, частоты пульса и др.

1.3. Методы вычисления коэффициента корреляции.

Для вычисления коэффициента корреляции используют методы рангов, или метод Спирмена «ро» (ρ), квадратов, или метод Пирсона (r), корреляционной решетки (η) и множественной корреляции. Наиболее простым методом является вычисление коэффициента корреляции методом рангов (метод Спирмена), но полученный метод дает приближенные результаты.

1.4. Коэффициент ранговой корреляции Спирмена.

Коэффициент ранговой корреляции Спирмена – это непараметрический метод, который используется с целью статистического изучения связи между явлениями. В этом случае определяется фактическая степень параллелизма между двумя количественными рядами изучаемых признаков и дается оценка тесноты установленной связи с помощью количественно выраженного коэффициента.

Практический расчет коэффициента ранговой корреляции Спирмена включает следующие этапы:

- составить два ряда из парных сопоставляемых признаков, обозначив первый и второй ряд соответственно X и Y. При этом представить первый ряд признака в убывающем или возрастающем порядке, а числовые значения второго ряда расположить напротив тех значений первого ряда, которым они соответствуют;
- величину признака в каждом из сравниваемых рядов заменить порядковым номером (рангом). Рангами, или номерами, обозначают места показателей (значения) первого и второго рядов. При этом числовым значениям второго признака ранги должны присваиваться в том же порядке, какой был принят при раздаче их величинам первого признака. При одинаковых величинах признака в ряду ранги следует определять как среднее число из суммы порядковых номеров этих величин;

- определить разность рангов между x и y (d): $d = x - y$;
- возвести полученную разность рангов в квадрат (d^2);
- получить сумму квадратов разности (Σd^2) и подставить полученные значения в формулу:

$$\rho_{xy} = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

где Σd^2 – сумма квадратов разностей рангов;
 1 и 6 – постоянные коэффициенты;
 n – число наблюдений сравниваемых пар.

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее, показателями слабой тесноты связи; значения более 0,4, но менее 0,7 – показателями умеренной тесноты связи, а значения 0,7 и более – показателями высокой тесноты связи.

Коэффициент ранговой корреляции целесообразно применять:

- при наличии небольшого количества наблюдений,
- когда нет необходимости в точном установлении силы связи, а достаточно ориентировочных данных,
- когда признаки представлены не только количественными, но и атрибутивными значениями,
- когда ряды распределения признаков имеют открытые варианты (например, стаж работы до 1 года, 20 лет и более и др.).

Мощность коэффициента ранговой корреляции Спирмена несколько уступает мощности параметрического коэффициента корреляции.

Наиболее точным и часто применяемым является метод квадратов, или метод Пирсона.

1.5. Метод квадратов (метод Пирсона).

Метод квадратов (метод Пирсона) применяется:

- когда требуется точное установление силы связи между признаками;
- когда признаки имеют только количественное выражение.

Коэффициент корреляции показывает степень статистической зависимости между двумя числовыми переменными. Он вычисляется следующим образом:

$$r_{xy} = \frac{\Sigma(d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}}$$

где r – коэффициент корреляции, вычисленный методом квадратов,
 d_x – отклонения вариантов от средней величины ($V_x - M_x$),
 d_y – отклонения вариантов от средней величины ($V_y - M_y$).

Практический расчет коэффициента корреляции (метод Пирсона) включает следующие этапы:

- построить вариационные ряды для каждого из сопоставляемых признаков, обозначив первый и второй ряд чисел соответственно x и y ;
- определить для каждого вариационного ряда средние значения (M_1 и M_2);
- найти отклонения (d_x и d_y) каждого числового значения от среднего значения своего вариационного ряда;
- полученные отклонения перемножить ($d_x \times d_y$);
- каждое отклонение возвести в квадрат и суммировать по каждому ряду (Σd_x^2 и d_y^2);
- подставить полученные значения в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\Sigma(d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}}$$

Достоверность коэффициента корреляции определяется величиной ошибки и доверительным коэффициентом t . В том случае, если полученный коэффициент корреляции в 3 раза и более превышает свою ошибку, он считается достоверным.

1.6. Вычисление ошибки коэффициента корреляции.

1. Ошибка коэффициента корреляции, вычисленного ранговым методом (Спирмена):

$$m_{\rho xy} = \sqrt{\frac{1 - \rho_{xy}^2}{n - 2}}$$

где m_{ρ} – средняя ошибка коэффициента корреляции, вычисленного методом рангов;
 ρ – величина коэффициента корреляции, вычисленного методом рангов;
 n – число наблюдений.

2. Ошибка коэффициента корреляции, вычисленного методом квадратов (Пирсона):

$$m_{rxy} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}$$

где m_r – ошибка коэффициента корреляции, вычисленного методом квадратов;
 n – число наблюдений.

Оценка достоверности коэффициента корреляции, полученного методом ранговой корреляции и методом квадратов:

Способ 1.

Достоверность определяется по формуле:

$$t = \frac{\rho_{xy}}{m_{\rho xy}} \quad \text{или} \quad t = \frac{r_{xy}}{m_{rxy}}$$

Критерий t оценивается по таблице значений t с учетом числа степеней свободы ($n - 2$), где n – число парных вариантов. Критерий t должен быть равен или больше табличного, соответствующего вероятности $p \geq 99\%$.

Способ 2.

Достоверность оценивается по специальной таблице стандартных коэффициентов корреляции. При этом достоверным считается такой коэффициент корреляции, когда при определенном числе степеней свободы ($n - 2$), он равен или более табличного, соответствующего степени безошибочного прогноза $p \geq 95\%$.

Контрольные вопросы

1. Корреляция, определение.
2. Виды проявления количественных связей между признаками.
3. Определения функциональной и корреляционной связи.
4. Практическое значение установления корреляционной связи.
5. Коэффициент корреляции.
6. Направление корреляционной связи.
7. Сила корреляционной связи.
8. Методы определения коэффициента корреляции и формулы.
9. Рекомендации по применению метода ранговой корреляции (метод Спирмена).
10. Рекомендации к применению метода квадратов (метод Пирсона).

ТИПОВЫЕ ЗАДАЧИ.

Задача 1 – эталон

на применение рангового метода

Задание: методом рангов установить направление и силу связи между стажем работы в годах и частотой травм, если получены следующие данные (табл.1):

Таблица 1.

Стаж работы в годах	Число травм на 100 работающих
до 1 года	24
1-2	16
3-4	12
5-6	12
7 и более	6

Обоснование выбора метода: для решения задачи может быть выбран только метод ранговой корреляции, т.к. первый ряд признака «стаж работы в годах» имеет открытые варианты (стаж работы до 1 года и 7 и более лет), что не позволяет использовать для установления связи между сопоставляемыми признаками более точный метод – метод квадратов.

Решение. Последовательность расчетов изложена в тексте, результаты представлены в табл. 2.

Таблица 2.

Стаж работы в годах	Число травм	Порядковые номера (ранги)		Разность рангов	Квадрат разности рангов
		X	Y	$d = (x - y)$	d^2
До 1 года	24	1	5	-4	16
1-2	16	2	4	-2	4
3-4	12	3	2,5	+0,5	0,25
5-6	12	4	2,5	+1,5	2,25
7 и более	6	5	1	+4	16
					$\Sigma d^2 = 38,5$

1. Каждый из рядов парных признаков обозначить через "х" и через "у" (графы 1-2).

2. Величину каждого из признаков заменить ранговым (порядковым) номером. Порядок раздачи рангов в ряду "х" следующий: минимальному значению признака (стаж до 1 года) присвоен порядковый номер "1", последующим вариантам этого же ряда признака соответственно в порядке увеличения 2-й, 3-й, 4-й и 5-й порядковые номера – ранги (см. графу 3).

Аналогичный порядок соблюдается при раздаче рангов второму признаку "у" (графа 4).

В тех случаях, когда встречаются несколько одинаковых по величине вариантов (например, в задаче-эталоне это 12 и 12 травм на 100 работающих при стаже 3-4 года и 5-6 лет, порядковый номер обозначить средним числом из суммы их порядковых номеров. Эти данные о числе травм (12 травм) при ранжировании должны занимать 2 и 3 места, среднее число из них равно $\frac{2+3}{2} = 2,5$.

Таким образом, числу травм "12" и "12" (признаку) следует раздать ранговые номера одинаковые – "2,5" (графа 4).

3. Определить разность рангов $d = (x - y)$ – (графа 5).

4. Разность рангов возвести в квадрат (d^2) и получить сумму квадратов разности рангов Σd^2 (графа 6).

5. Произвести расчет коэффициента ранговой корреляции по формуле:

$$\rho_{xy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)},$$

где n – число сопоставляемых пар вариант в ряду "х" и в ряду "у"

$$\rho_{xy} = 1 - \frac{6 \times 38,5}{5(5^2 - 1)} = 1 - \frac{325}{5(25 - 1)} = 1 - \frac{325}{120} = 1 - 2,71 = -0,71$$

Определить достоверность коэффициента ранговой корреляции:

1-й способ. Определить ошибку ($m_{\rho_{xy}}$) коэффициента ранговой корреляции и оценить достоверность его с помощью критерия t:

$$m_{\rho_{xy}} = \sqrt{\frac{1 - \rho_{xy}^2}{n - 2}} = \frac{1 - 0,71^2}{n - 2} = \sqrt{\frac{0,08}{3}} = \sqrt{0,026} = 0,16$$

$$t = \frac{\rho_{xy}}{m_{\rho_{xy}}} = \frac{0,71}{0,16} = 4,44$$

Полученный критерий $t = 4,44$ соответствует вероятности безошибочного прогноза p больше 95%:

$$\rho_{xy} = -0,71; m_{\rho_{xy}} = \pm 0,16; t = 4,44; p > 95\%.$$

2-й способ. По таблице "Стандартных коэффициентов корреляции": при числе степеней свободы $(n - 2) = 5 - 2 = 3$ наш расчетный коэффициент корреляции $\rho_{xy} = -0,71$ больше табличного 0,878 и меньше 0,934, что соответствует вероятности безошибочного прогноза больше 95% и меньше 98%. Это позволяет считать полученный коэффициент ранговой корреляции достоверным.

Вывод. С вероятностью безошибочного прогноза (p) больше 95% установлено, что чем больше стаж работы, тем меньше частота травм (связь обратная, сильная, достоверная корреляционная: $\rho_{xy} = -0,71, p > 95\%$).

Задача 2 – эталон

на применение метода квадратов (метод Пирсона).

Задание: вычислить коэффициент корреляции, определить направление и силу связи между количеством кальция в воде и жесткостью воды, если известны следующие данные (табл. 1). Оценить достоверность связи. Сделать вывод.

Таблица 1.

Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262

Обоснование выбора метода. Для решения задачи выбран метод квадратов (Пирсона), т.к. каждый из признаков (жесткость воды и количество кальция) имеет числовое выражение; нет открытых вариантов.

Решение. Последовательность расчетов изложена в тексте, результаты представлены в таблице 2. Построив ряды из парных сопоставляемых признаков, обозначить их через x (жесткость воды в градусах) и через y (количество кальция в воде в мг/л).

Таблица 2.

Жесткость воды (в градусах) x	Количество кальция в воде (в мг/л) y	d _x	d _y	d _x x d _y	d _x ²	d _y ²
4	28	-16	-114	1824	256	12996
8	56	-12	-86	1032	144	7396
11	77	-9	-66	594	81	4356
27	191	+7	+48	336	49	2304
34	241	+14	+98	1372	196	9604
37	262	+16	+120	1920	256	14400
M _x = Σ x / n	M _y = Σ y / n			Σd _x x d _y = 7078	Σd _x ² = 982	Σd _y ² = 51056
M _x = 120/6 = 20	M _y = 852/6 = 142					

1. Определить средние величины M_x ряду вариант "x" и M_y в ряду вариант "y" по формулам:

$$M_x = \Sigma x / n \text{ (графа 1),}$$

$$M_y = \Sigma y / n \text{ (графа 2);}$$

2. Найти отклонение (d_x и d_y) каждой варианты от величины вычисленной средней в ряду "x" и в ряду "y":

$$d_x = x - M_x \text{ (графа 3) и } d_y = y - M_y \text{ (графа 4);}$$

3. Найти произведение отклонений d_x x d_y и суммировать их: Σ d_x x d_y (графа 5);

4. Каждое отклонение d_x и d_y возвести в квадрат, и суммировать их значения по ряду "x" и по ряду "y": Σ d_x² = 982 (графа 6) и Σd_y² = 51056 (графа 7);

5. Определить произведение Σd_x² x Σ d_y² и из этого произведения извлечь квадратный корень:

$$\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)} = \sqrt{(982 \times 51056)}$$

6. Полученные величины Σ (d_x x d_y) и √(Σd_x² x Σd_y²) подставляем в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\Sigma(d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}} = \frac{7078}{\sqrt{(982 \times 51056)}} = \frac{7078}{\sqrt{50136992}} = \frac{7078}{7080.7} = +0.99$$

7. Определить достоверность коэффициента корреляции:

1-й способ. Найти ошибку коэффициента корреляции (m_{r_{xy}}) и критерий t по формулам:

$$m_{r_{xy}} = \pm \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \pm \sqrt{\frac{1 - 0,99^2}{6 - 2}} = \pm \sqrt{\frac{0,02}{4}} = \pm \sqrt{0,005} = \pm 0,07$$

$$t = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,99}{0,07} = 14,1$$

Критерий t = 14,1, что соответствует вероятности безошибочного прогноза ρ > 99,9%.

2-й способ. Достоверность коэффициента корреляции оценивается по таблице "Стандартные коэффициенты корреляции" (см. приложение 1). При числе степеней свободы (n - 2) = 6 - 2 = 4, наш расчетный коэффициент корреляции r_{xy} = + 0,99 больше табличного (r_{табл} = + 0,917 при ρ = 99%).

Вывод. Чем больше кальция в воде, тем она более жесткая (связь прямая, сильная и достоверная: r_{xy} = + 0,99, ρ > 99,9%).

Задача 3 – эталон.

Определить коэффициент корреляции Пирсона, сделать выводы. Результаты определения потребного количества калорий в сутки в зависимости от веса, у девочек 8 лет (табл. 1).

Таблица 1.

Средний вес 8-летних девочек (кг)	Потребное количество ккал.в сутки/кг веса
21	61,9
22	63,6
24	62,5
25	64,0
27	61,1
28	60,7
29	60,3

Построив ряды из парных сопоставляемых признаков, обозначить их через x (вес девочек в килограммах) и через y (потребного количества килокалории в сутки в зависимости от веса, у девочек 8 лет) (табл. 2)

Таблица 2.

	Вес девочек (кг), X	Потребное количество ккал в сутки/кг веса, Y	$d_x =$	d_y	$d_x \times d_y$	d_x^2	d_y^2
1	21	61,9	4	4,1	0,1	0,41	16,81
2	22	63,6	6	3,1	-1,6	-4,96	9,61
3	24	62,5	5	1,1	-0,5	-0,55	1,21
4	25	64,0	7	0,1	-2,0	-0,2	0,01
5	27	61,1	3	-1,9	0,9	-1,71	3,61
6	28	60,7	2	-2,9	1,3	-3,77	8,41
7	29	60,3	1	-3,9	1,7	-6,63	15,21
	$M_x = \Sigma x / n$	$M_y = \Sigma y / n$			$\Sigma d_x \times d_y$ = - 17,41	$\Sigma d_x^2 =$ 54,87	$\Sigma d_y^2 =$ 12,21
	$M_x = 176 / 7 =$ 25,1	$M_y = 434,1 / 7$ = 62,0					

1. Определить средние величины M_x ряду вариант "x" и M_y в ряду вариант "y" по формулам:
 $M_x = \Sigma x / n$ (графа 1),
 $M_y = \Sigma y / n$ (графа 2);
2. Найти отклонение (d_x и d_y) каждой варианты от величины вычисленной средней в ряду "x" и в ряду "y":
 $d_x = x - M_x$ (графа 3) и $d_y = y - M_y$ (графа 4);
3. Найти произведение отклонений $d_x \times d_y$ и суммировать их: $\Sigma d_x \times d_y$ (графа 5);
4. Каждое отклонение d_x и d_y возвести в квадрат, и суммировать их значения по ряду "x" и по ряду "y": $\Sigma d_x^2 = 982$ (графа 6) и $\Sigma d_y^2 = 51056$ (графа 7);
5. Определить произведение $\Sigma d_x^2 \times \Sigma d_y^2$ и из этого произведения извлечь квадратный корень:

$$\sqrt{\Sigma d_x^2 \times \Sigma d_y^2} = \sqrt{(54,87 \times 12,21)}$$

6. Полученные величины $\Sigma (d_x \times d_y)$ и $\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}$ подставляем в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\Sigma(d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}} = \frac{-17,41}{\sqrt{(54,87 \times 12,21)}} = \frac{-17,41}{\sqrt{669,96}} = \frac{-17,41}{25,88} = -0,67$$

Вывод: связь корреляционная, сильная, обратная.

7. Определить достоверность коэффициента корреляции: Представительность коэффициента корреляции может быть определена по его средней ошибке, которую можно вычислить по формуле:

$$m_{r_{xy}} = \pm \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \pm \sqrt{\frac{1 - (-0,67)^2}{7 - 2}} = \pm \sqrt{\frac{1 - 0,45}{5}} = \pm \sqrt{0,11} = \pm 0,33$$

Коэффициент корреляции достоверен в том случае, если он превышает свою ошибку в 3 и более раз. Это условие в данном случае не выполняется.

8. Оценка значимости по t-критерию:

$$t = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,67}{0,33} = 2,03$$

Так как $t > 2$, то это говорит о значимости критерия.

Вывод: в результате проведенного анализа полученных данных можно говорить о наличии сильной и обратной взаимосвязи между средним весом 8-летних девочек и потребным количеством калорий в сутки. При расчете и оценке коэффициента Пирсона вычислили его ошибку: т.к. она не более чем в 3 раза меньше самого коэффициента, можно говорить о недостоверности данного коэффициента. Также была проведена оценка значимости по t-критерию: т.к. $t > 2$, можно говорить о значимости полученных результатов. Таким образом, можно говорить о недостаточной достоверности влияния веса 8-летних девочек на потребное количество калорий в сутки и о значимости полученных результатов.

Задача 4 – эталон.

На основе приведенных данных в таблице 1 требуется: 1) вычислить коэффициент корреляции рангов; 2) определить характер и силу связи между соответствующими признаками; определить достоверность коэффициента корреляции.

Задание: длина и масса тела у 12 девочек в возрасте 5 лет (табл. 1):

Таблица 1.

Длина тела (см.)	Масса тела (кг.)
87	13
95	14
115	20
89	12
90	14
90	15
101	17
95	15
110	18
110	21
88	14
93	16

Решение.

1. Рангами (порядковыми номерами) обозначаем места показателей в рядах «х» и «у» (табл. 2), затем находим разность между рангами (d) и возводим ее в квадрат (d^2). При обозначении места показателей рангами начинают с меньшего (или с большего).

Если отдельные показатели ряда встречаются несколько раз (14, 15, 90, 95, 110) ранги проставляются следующим образом: масса тела 14 кг. встречается трижды занимая по величине 3-е, 4-е и 5-е места, поэтому порядковые номера в этом случае будут равны $\frac{3+4+5}{3} = 4$, т.е. против каждого показателя 14 кг. будет проставлен ранг 4 и т.д.

Таблица 2.

Длина тела (см.) x	Масса тела (кг.) y	Порядковые номера (ранги)		Разность рангов d = (x – y)	Квадрат разности рангов d ²
		X	Y		
87	13	1	2	-1	1
95	14	7,5	4	+3,5	12,25
115	20	12	11	+1	1
89	12	3	1	+2	4
90	14	4,5	4	+0,5	0,25
90	15	4,5	6,5	-2	4
101	17	9	9	0	0
95	15	7,5	6,5	+1	1
110	18	10,5	10	+0,5	0,25
110	21	10,5	12	-1,5	2,25
88	14	2	4	-2	4
93	16	6	8	-2	4
					Σ d ² = 34

Подставляем полученные данные в формулу коэффициента корреляции рангов:

$$\rho_{xy} = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 34}{12(12^2-1)} = 1 - \frac{204}{1716} = 1 - 0,12 = 0,881$$

2. Коэффициент корреляции, равный + 0,881, свидетельствует о наличии прямой сильной связи между ростом девочек и массой их тела.

3. Определяем достоверность коэффициента ранговой корреляции:

1-й способ. Определяем ошибку ($m_{\rho_{xy}}$) коэффициента ранговой корреляции и оцениваем достоверность его с помощью критерия t:

$$m_{\rho_{xy}} = \sqrt{\frac{1-\rho_{xy}^2}{n-2}} = \sqrt{\frac{1-0,881^2}{12-2}} = \sqrt{\frac{1-0,776}{10}} = \sqrt{\frac{0,224}{10}} = \sqrt{0,0224} = 0,15$$

$$t = \frac{\rho_{xy}}{m_{\rho_{xy}}} = \frac{0,881}{0,15} = 5,867$$

Полученный критерий $t = 5,867$ соответствует вероятности безошибочного прогноза ρ больше 95%: $\rho_{xy} = 0,881$; $m_{\rho_{xy}} = \pm 0,15$; $t = 5,867$; $\rho > 95\%$.

2-й способ. Достоверность коэффициента корреляции оценивается по таблице «Стандартных коэффициентов корреляции». При числе степеней свободы $(n - 2) = 12 - 2 = 10$ наш расчетный коэффициент корреляции $\rho_{xy} = 0,881$ больше табличного 0,708 что соответствует вероятности безошибочного прогноза больше 99%. Это позволяет считать полученный коэффициент ранговой корреляции достоверным.

Вывод. С вероятностью безошибочного прогноза (ρ) больше 99% установлено, что чем больше рост, тем больше масса тела (связь прямая, сильная, достоверная корреляционная: $\rho_{xy} = 0,881$, $\rho > 99\%$).

Задача 5 – эталон.

Путем вычисления коэффициента ранговой корреляции Спирмена определить характер и размер связи между уровнем мертворождаемости и весом ребенка при рождении (табл. 1).

Таблица 1.

Вес в граммах	Мертворождаемость (на 1000 родов)
до 1500	137,2
1500 – 1749	86
1750 – 1999	44,2
2000 – 2249	35,6
2250 – 2499	18
2500 – 2749	12
2750 – 3000	7,2

Решение:

1. Каждый из рядов парных признаков обозначить через "х" и через "у" (графы 1-2).

2. Величину каждого из признаков заменить ранговым (порядковым) номером. Порядок раздачи рангов в ряду "х" следующий: минимальному значению признака (вес в граммах до 1500) присвоен порядковый номер "1", последующим вариантам этого же ряда признака соответственно в порядке увеличения 2-й, 3-й, 4-й и 5-й порядковые номера – ранги (см. графу 3).

Аналогичный порядок соблюдается при раздаче рангов второму признаку "у" (графа 4) (табл. 2).

Таблица 2.

Вес (в граммах), X	Мертворождаемость, Y	Ранги		d	d ²
		X	Y		
до 1500	137,2	1	7	– 6	36
1500 – 1749	86	2	6	– 4	16
1750 – 1999	44,2	3	5	– 2	4
2000 – 2249	35,6	4	4	0	0
2250 – 2499	18	5	3	2	4
2500 – 2749	12	6	2	4	16
2750 – 3000	7,2	7	1	6	36
n = 7					$\Sigma d^2 = 112$

3. Определить разность рангов $d = (x - y)$ – (графа 5).

4. Разность рангов возвести в квадрат (d^2) и получить сумму квадратов разности рангов Σd^2 (графа 6).

5. Произвести расчет коэффициента ранговой корреляции по формуле:

$$\rho_{xy} = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 112}{7(7^2 - 1)} = 1 - \frac{672}{336} = 1 - 2 = -1$$

где n – число сопоставляемых пар вариант в ряду "х" и в ряду "у"

Вывод: установлена функциональная связь.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

ЗАДАНИЕ: ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ РАНГОВ И ОЦЕНКА ЕГО ДОСТОВЕРНОСТИ.

На основе приведенных данных требуется: 1) вычислить коэффициент корреляции рангов; 2) определить характер и силу связи между соответствующими признаками; 3) определить достоверность коэффициента корреляции.

Вариант 1

При проведении комплексных медицинских осмотров у лиц разных возрастов число хронических заболеваний (на 1000 осмотренных данного возраста составило (табл. 1):

Таблица 1

Возраст, годы	Число хронических заболеваний
0-4	748.6
5-9	903.8
10-14	982.4
15-19	1010.6
20-24	1281.6
25-29	1340.9
30-39	1679.6
40-49	1944.8
50-59	2635.8
60-69	3564.7
70-79	4071.8

Вариант 2

Результаты измерения роста и массы тела студентов в возрасте 20 лет:

Рост, см.	157	158	160	165	167	162	171	174	168	176	170	180
Масса тела, кг.	56	55	57	57	58	60	63	65	67	72	79	82

Вариант 3

Уровни систолического и диастолического давления (в мм.рт. ст.) у 12 здоровых юношей в возрасте 18 лет:

Систолическое	105	115	115	110	110	120	120	120	125	110	125	120
Диастолическое	65	70	65	65	70	75	75	70	75	70	80	80

Вариант 4

Длина и масса тела у 10 новорожденных:

Длина тела, см.	35	48	52	50	47	53	52	50	51	54
Масса тела, кг.	4,5	3,6	4,1	4,0	3,2	3,8	3,9	3,9	4,0	4,3

Вариант 5

Длина и масса тела у 10 девочек в возрасте 6 лет:

длина тела, см.	106	110	114	120	122	126
масса тела, кг.	18	19	21	22	22	24

ПРИЛОЖЕНИЕ 1.

Стандартные коэффициенты корреляции, которые считаются достоверными
(по Л.С. Каминскому)

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

ПРИЛОЖЕНИЕ 2.

Таблица значений критерия Стьюдента (t-критерия)

Критические значения коэффициента Стьюдента (t-критерия) для различной доверительной вероятности p и числа степеней свободы f :

f	p							
	0.80	0.90	0.95	0.98	0.99	0.995	0.998	0.999
1	3.0770	6.3130	12.7060	31.820	63.656	127.656	318.306	636.619
2	1.8850	2.9200	4.3020	6.964	9.924	14.089	22.327	31.599
3	1.6377	2.35340	3.182	4.540	5.840	7.458	10.214	12.924
4	1.5332	2.13180	2.776	3.746	4.604	5.597	7.173	8.610
5	1.4759	2.01500	2.570	3.649	4.0321	4.773	5.893	6.863
6	1.4390	1.943	2.4460	3.1420	3.7070	4.316	5.2070	5.958
7	1.4149	1.8946	2.3646	2.998	3.4995	4.2293	4.785	5.4079
8	1.3968	1.8596	2.3060	2.8965	3.3554	3.832	4.5008	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	3.6897	4.2968	4.780
10	1.3720	1.8125	2.2281	2.7638	3.1693	3.5814	4.1437	4.5869
11	1.363	1.795	2.201	2.718	3.105	3.496	4.024	4.437
12	1.3562	1.7823	2.1788	2.6810	3.0845	3.4284	3.929	4.178
13	1.3502	1.7709	2.1604	2.6503	3.1123	3.3725	3.852	4.220
14	1.3450	1.7613	2.1448	2.6245	2.976	3.3257	3.787	4.140

15	1.3406	1.7530	2.1314	2.6025	2.9467	3.2860	3.732	4.072
16	1.3360	1.7450	2.1190	2.5830	2.9200	3.2520	3.6860	4.0150
17	1.3334	1.7396	2.1098	2.5668	2.8982	3.2224	3.6458	3.965
18	1.3304	1.7341	2.1009	2.5514	2.8784	3.1966	3.6105	3.9216
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834
20	1.3253	1.7247	2.08600	2.5280	2.8453	3.1534	3.5518	3.8495
21	1.3230	1.7200	2.2.0790	2.5170	2.8310	3.1350	3.5270	3.8190
22	1.3212	1.7117	2.0739	2.5083	2.8188	3.1188	3.5050	3.7921
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.1040	3.4850	3.7676
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.0905	3.4668	3.7454
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782	3.4502	3.7251
26	1.315	1.705	2.059	2.478	2.778	3.0660	3.4360	3.7060
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565	3.4210	3.6896
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.0469	3.4082	3.6739
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.0360	3.3962	3.8494
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.0298	3.3852	3.6460
32	1.3080	1.6930	2.0360	2.4480	2.7380	3.0140	3.3650	3.6210
34	1.3070	1.6909	2.0322	2.4411	2.7284	3.9520	3.3479	3.6007
36	1.3050	1.6883	2.0281	2.4345	2.7195	9.490	3.3326	3.5821
38	1.3042	1.6860	2.0244	2.4286	2.7116	3.9808	3.3190	3.5657
40	1.303	1.6839	2.0211	2.4233	2.7045	3.9712	3.3069	3.5510
42	1.320	1.682	2.018	2.418	2.6980	2.6930	3.2960	3.5370
44	1.301	1.6802	2.0154	2.4141	2.6923	3.9555	3.2861	3.5258
46	1.300	1.6767	2.0129	2.4102	2.6870	3.9488	3.2771	3.5150
48	1.299	1.6772	2.0106	2.4056	2.6822	3.9426	3.2689	3.5051
50	1.298	1.6759	2.0086	2.4033	2.6778	3.9370	3.2614	3.4060
55	1.2997	1.673	2.0040	2.3960	2.6680	2.9240	3.2560	3.4760
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.9146	3.2317	3.4602
65	1.2947	1.6686	1.997	2.3851	2.6536	3.9060	3.2204	3.4466
70	1.2938	1.6689	1.9944	2.3808	2.6479	3.8987	3.2108	3.4350
80	1.2820	1.6640	1.9900	2.3730	2.6380	2.8870	3.1950	3.4160
90	1.2910	1.6620	1.9867	2.3885	2.6316	2.8779	3.1833	3.4019
100	1.2901	1.6602	1.9840	2.3642	2.6259	2.8707	3.1737	3.3905
120	1.2888	1.6577	1.9719	2.3578	2.6174	2.8598	3.1595	3.3735
150	1.2872	1.6551	1.9759	2.3515	2.6090	2.8482	3.1455	3.3566
200	1.2858	1.6525	1.9719	2.3451	2.6006	2.8385	3.1315	3.3398
250	1.2849	1.6510	1.9695	2.3414	2.5966	2.8222	3.1232	3.3299
300	1.2844	1.6499	1.9679	2.3388	2.5923	2.8279	3.1176	3.3233
400	1.2837	1.6487	1.9659	2.3357	2.5882	2.8227	3.1107	3.3150
500	1.2830	1.6470	1.9640	2.3330	2.7850	2.8190	3.1060	3.3100

Глава 2. СТАНДАРТИЗАЦИЯ.

2.1. Стандартизация, понятие. Методы стандартизации.

Общие интенсивные коэффициенты (рождаемости, смертности, детской смертности, заболеваемости и т.д.) правильно отражают частоту явлений при их сопоставлении лишь в том случае, если состав сравниваемых совокупностей однороден. Если же они имеют неоднородный возрастно-половой или профессиональный состав, различие по тяжести болезни, по нозологическим формам или по другим признакам, то ориентируясь на общие показатели, сравнивая их, можно сделать неправильный вывод о тенденциях изучаемых явлений и истинных причинах разницы общих показателей сравниваемых совокупностей. Даже при наличии одинаковой тенденции во всех сравниваемых группах не всегда удобно пользоваться набором показателей, а предпочтительнее получить единую суммарную оценку. Во всех подобных случаях прибегают к методу стандартизации, то есть к устранению (элиминации) влияния состава (структуры) совокупностей на общий, итоговый показатель.

Следовательно, метод стандартизации применяется тогда, когда имеющиеся различия в составе сравниваемых совокупностей могут повлиять на размеры общих коэффициентов.

Для того, чтобы устранить влияние неоднородности составов сравниваемых совокупностей на величину получаемых коэффициентов, их приводят к единому стандарту, то есть условно допускается, что состав сравниваемых совокупностей одинаков. В качестве стандарта можно принять состав какой-либо близкой по существу третьей совокупности, средний состав двух сравниваемых групп или, проще всего, состав одной из сравниваемых групп.

Стандартизация – метод сравнения показателей в двух неоднородных совокупностях на основании расчета условных (стандартизованных показателей при использовании стандарта).

При сравнении двух неоднородных совокупностей по какому-либо признаку (составу) применяются методы стандартизации (прямой, обратный, косвенный).

Прямой способ применяют, когда имеются погрупповые (повозрастные) показатели заболеваемости (смертности, травматизма) или их можно вычислить (при наличии погрупповой численности населения и заболевших).

Косвенный способ используют, если показатели по группам отсутствуют и их нельзя вычислить из-за отсутствия числа заболевших.

Обратный способ применяют при отсутствии погрупповых величин численности населения.

Наиболее распространенным является прямой метод стандартизации.

2.2. Прямой метод стандартизации.

Прямой метод применяется:

- при наличии полных сведений, как о составе сравниваемых совокупностей, так и о распределении в них явления,
- при сравнении интенсивных показателей в совокупностях, отличающихся по составу (например, по возрасту, полу, профессиям и т.д.).

Сущность метода стандартизации. Он позволяет устранить (элиминировать) возможное влияние различий в составе совокупностей по какому-либо признаку на величину сравниваемых интенсивных показателей. С этой целью составы совокупностей по данному признаку уравниваются, что в дальнейшем позволяет рассчитать стандартизованные показатели.

Метод стандартизации используется при оценке показателей здоровья только при сравнении их уровней. Этот метод расчета условных величин применяется для устранения неоднородности состава сравниваемых коллективов. Он показывает, какой был бы уровень

заболеваемости (травматизма, смертности, инвалидизации и др.) в каждом коллективе (учреждении, городе), если бы его состав (по возрасту, по полу, по стажу и др.) был одинаков.

Стандартизованные показатели – это условные, гипотетические величины, они не отражают истинных размеров явлений. Стандартизованные показатели свидетельствуют о том, каковы были бы значения сравниваемых интенсивных показателей, если бы были исключены различия в составах совокупностей.

Назначение метода стандартизации. Метод стандартизации применяется для выявления влияния фактора неоднородности составов совокупностей по какому-либо признаку на различия сравниваемых интенсивных показателей.

2.3. Этапы расчета стандартизованных показателей.

I этап. Расчет общих и частных интенсивных показателей:

- *общих* – по совокупностям в целом;
- *частных* – по признаку различия (полу, возрасту, стажу работы и т.д.).

II этап. Определение стандарта, т.е. выбор одинакового численного состава среды по данному признаку (по возрасту, полу и т.д.) для сравниваемых совокупностей. Как правило, за стандарт принимается сумма или полусумма численностей составов соответствующих групп. В то же время стандартом может стать состав любой из сравниваемых совокупностей, а также состав по аналогичному признаку какой-либо другой совокупности. Например, при сравнении летальности в конкретной больнице по двум отделениям скорой помощи за стандарт может быть выбран состав больных любой другой больницы скорой помощи. Таким образом, так или иначе уравниваются условия среды, что дает возможность провести расчеты новых чисел явления, называемых "ожидаемыми величинами".

III этап. Вычисление ожидаемых абсолютных величин в группах стандарта на основе групповых интенсивных показателей, рассчитанных на I этапе. Итоговые числа по сравниваемым совокупностям являются суммой ожидаемых величин в группах.

IV этап. Вычисление стандартизованных показателей для сравниваемых совокупностей.

V этап. Сопоставление соотношений стандартизованных и интенсивных показателей, формулировка вывода.

Общим этапом вычисления стандартизованных коэффициентов является выбор стандарта возрастно-полового состава (процентное распределение состава любой из сравниваемых групп или их суммарного значения). При выборе стандартного состава уровня заболеваемости можно использовать литературные данные или показатели предыдущих исследований.

Контрольные вопросы:

1. Стандартизация, определение.
2. Для чего применяется стандартизация?
3. Методы стандартизации.
4. Прямой метод стандартизации.
5. Этапы прямого метода стандартизации
6. Косвенный метод стандартизации.
7. Обратный метод стандартизации.
8. О чем свидетельствуют стандартизованные показатели?

ТИПОВЫЕ ЗАДАЧИ.

Вычисление стандартизованных показателей

На основе приведенных в табл.1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.
2. Сравнить:
 - общие показатели, рассчитанные обычным способом;
 - частные показатели в отдельных группах;
 - стандартизованные и обычные показатели.
3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Задача 1 – эталон.

Задание. Используя метод стандартизации при сравнении уровней летальности в больницах А и Б (табл. 1), сделайте соответствующие выводы.

Таблица 1.

Возраст больных (в годах)	Больница А		Больница Б	
	число выбывших больных	из них умерло	число выбывших больных	из них умерло
До 40	600	12	1400	42
От 40 до 59	200	8	200	10
От 60 и старше	1200	60	400	24
Всего:	2000	80	2000	76

Решение:

Этапы расчета стандартизованных показателей:

I этап. Сначала определяют общие показатели летальности в больницах А и Б.

Больница А: $\frac{80 \times 100}{2000} = 4$ на 100 выбывших больных;

Больница Б: $\frac{76 \times 100}{2000} = 3,8$ на 100 выбывших больных.

Затем находят показатели летальности в зависимости от возраста больных. Например: в больнице А у больных в возрасте до 40 лет летальность составляет $\frac{12 \times 100}{600} = 2$ на 100 выбывших больных, а в больнице Б, соответственно, $\frac{42 \times 100}{1400} = 3$ на 100 выбывших больных (табл. 2).

Аналогично проводят расчеты и в других возрастных группах (см. сводную табл. 2 – I этап).

Таблица 2.

Показатели летальности в зависимости от возраста (Больница А и Б)

Возраст больных (в годах)	Больница А			Больница Б		
	выбывших больных	из них умерло	Показатель летальности по возрастам на 100 выбывших больных	число выбывших больных	из них умерло	Показатель летальности по возрастам на 100 выбывших больных
До 40	600	12	2	1400	42	3
От 40 до 59	200	8	4	200	10	5
От 60 и старше	1200	60	5	400	24	6
Всего:	2000	80	4	2000	76	3,8

II этап. За стандарт принимают сумму выбывших больных по каждой возрастной группе в обеих больницах (табл. 3).

Таблица 3.

Возраст больных (в годах)	Число больных в больницах А и Б	Стандарт
До 40	600 + 1400	2000
От 40 до 59	200 + 200	400
От 60 и старше	1200 + 400	1600
Всего:	2000 + 2000	4000

III этап. Определяют ожидаемое число умерших в стандарте по каждой возрастной группе в больницах А и Б, с учетом соответствующих показателей летальности (табл. 4):

Находят сумму ожидаемых чисел умерших в стандарте в больнице А ($40 + 16 + 80 = 136$) и больнице Б ($60 + 20 + 96 = 176$) (см. табл. 4).

Таблица 4.

Возраст до 40 лет:	Возраст от 40 до 59:	Возраст 60 лет и старше:
Больница А. $100 - 2$ $2000 - X$ $X = \frac{2 \times 2000}{100} = 40$ Больница Б. $100 - 3$ $2000 - X$ $X = \frac{3 \times 200}{100} = 60$	Больница А. $100 - 4$ $40 - X$ $X = \frac{4 \times 400}{100} = 16$ Больница Б. $100 - 5$ $400 - X$ $X = \frac{5 \times 400}{100} = 20$	Больница А. $100 - 5$ $1600 - X$ $X = \frac{5 \times 1600}{100} = 80$ Больница Б. $100 - 6$ $1600 - X$ $X = \frac{6 \times 1600}{100} = 96$

IV этап. Определяют общие стандартизованные показатели травматизма в больницах А и Б.

Больница А. $136 \times 100 / 4000 = 3,4$ на 100 выбывших больных;

Больница Б. $176 \times 100 / 4000 = 4,4$ на 100 выбывших больных.

Результаты поэтапного расчета стандартизованных показателей летальности оформляют в виде таблицы (табл. 5).

Таблица 5.

Результаты поэтапного расчета стандартизованных показателей летальности

Возраст больных (в годах)	Больница А		Больница Б		I этап		II этап	III этап	
	Выбыло больных	из них умерло	выбыло больных	из них умерло	летальность на 100 выбывших больных		Стандарт (сумма составов больных обеих больниц)	ожидаемое число умерших в стандарте	
					Б-ца А	Б-ца Б		Б-ца А	Б-ца Б
До 40 лет	600	12	1400	42	2	3	2000	40	60
От 40 до 59	200	8	200	10	4	5	400	16	20
60 и старше	1200	60	400	24	5	6	1600	80	96
Всего:	2000	80	2000	76	4,5	3,8	4000	136	176
IV этап. Определение стандартизованных показателей							100	3,4	4,4

V этап. Сопоставление соотношения интенсивных и стандартных показателей летальности в больницах А и Б (табл. 6).

Таблица 6.

Показатели:	Больница А	Больница Б	Соотношение А и Б
Интенсивные	4,0	3,8	$A > B$
Стандартные	3,4	4,4	$A < B$

Выводы:

1. Уровень летальности в больнице А выше, чем в больнице Б.
2. Однако если бы возрастной состав выбывших больных в этих больницах был одинаков, то летальность была бы выше в больнице Б.
3. Следовательно, на различия в уровнях летальности (в частности, на "завышение" ее в больнице А и "занижение" в больнице Б) оказала влияние неоднородность возрастного состава больных, а именно, преобладание в больнице А пожилых пациентов (60 лет и более) с относительно высоким показателем летальности, и наоборот, в больнице Б – больных в возрасте до 40 лет, имеющих низкие показатели летальности.

Задача 2 – эталон.

На основе приведенных в табл. 1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.
2. Сравнить:
 - общие показатели, рассчитанные обычным способом;
 - частные показатели в отдельных группах;
 - стандартизованные и обычные показатели.
3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Вариант 1.

Распределение лиц, имевших травмы на промышленном предприятии, по полу за два периода времени (в абс. числах) (табл. 1).

Таблица 1.

Пол	Первый период		Второй период	
	Число работающих	Число травм	Число работающих	Число травм
Мужчины	200	32	600	72
Женщины	400	28	200	16
Всего	600	60	800	88

Решение:

1. Расчет стандартизованных показателей травм на промышленном предприятии.

I этап. Этапы расчета стандартизованных показателей.

Сначала определяют общие показатели травм в первом и втором периодах времени:

первый период: $\frac{60 \times 100}{600} = 10$ на 100 работающих; второй период: $\frac{88 \times 100}{800} = 11$ на 100 работающих.

Затем находят показатели травматизма в зависимости от пола: среди мужчин травматизм в первом периоде составил $\frac{32 \times 100}{200} = 16$ на 100 работающих, втором периоде $\frac{72 \times 100}{600} = 12$ на 100 работающих; женщин (соответственно) $\frac{28 \times 100}{400} = 7$ и $\frac{16 \times 100}{200} = 8$ (см. сводную табл. 2).

Таблица 2.

Распределение лиц, имевших травмы на промышленном предприятии, по полу за два периода времени (на 100 работающих).

Пол	Первый период (ПП)			Второй период (ВП)		
	Число работающих (абс. ч.)	Число травм (абс. ч.)	Показатель травматизма на 100 работающих	Число работающих (абс. ч.)	Число травм (абс. ч.)	Показатель травматизма на 100 работающих
Мужчины	200	32	16	600	72	12
Женщины	400	28	7	200	16	8
Всего	600	60	10	800	88	11

II этап. Определение стандарта.

За стандарт принимают полусумму состава по полу за оба периода времени (табл. 3).

Таблица 3.

Пол	Первый период	Второй период	Число работающих за оба периода	Стандарт
	Число работающих	Число работающих		
Мужчины	200	600	$200 + 600 : 2$	400
Женщины	400	200	$400 + 200 : 2$	300
Всего	600	800	$600 + 800 : 2$	700

III этап. Расчет ожидаемых чисел (в данном случае травм) в каждой группе стандарта.

Если показатель травматизма на 100 работающих-мужчин на промышленном предприятии в первом периоде составляло 16 человек, то сколько бы составил данный показатель, если бы число работающих мужчин равнялся бы 400 (стандарт).

Составляют пропорцию $100 - 16$

$$400 - X \quad X = \frac{16 \times 400}{100} = 64$$

Из данной пропорции получаем ожидаемую величину. Аналогично полученные данные в целом представлены в таблице 4.

Находят сумму ожидаемых чисел травм в стандарте первого периода ($64 + 21 = 85$) и второго периода ($48 + 24 = 72$) (см. табл. 4).

Таблица 4.

Расчет ожидаемых величин травм в первом и втором периодах времени:

Мужчин:	Женщин:
Первый период $100 - 16$ $400 - X$ $X = \frac{16 \times 400}{100} = 64$	Первый период $100 - 7$ $300 - X$ $X = \frac{7 \times 300}{100} = 21$
Второй период $100 - 12$ $400 - X$ $X = \frac{12 \times 400}{100} = 48$	Второй период $100 - 8$ $300 - X$ $X = \frac{8 \times 300}{100} = 24$

IV этап. Расчет стандартизованных коэффициентов.

Определяют общие стандартизованные показатели травматизма за оба периода.

При условии, что в каждом периоде число работающих составляло 700 (стандарт, см. табл. 3), рассуждаем следующим образом. Из 700 работающих в первом периоде ожидаемое

число травм составляет 85, следовательно, показатель травматизма вычисляется на основе пропорции:

Первый период: $85 \times 100 / 700 = 12,1$ на 100 выбывших больных;

Второй период: $72 \times 100 / 700 = 10,3$ на 100 выбывших больных.

Это и есть стандартизованные показатели, т.е. показатели, вычисленные при условии, что состав работающих в каждом периоде одинаковый (стандартный).

Результаты поэтапного расчета стандартизованных показателей травматизма оформляют в виде таблицы (табл. 5).

Таблица 5.

Результаты поэтапного расчета стандартизованных показателей травматизма:

Пол	Первый период «А»		Второй период «Б»		I этап		II этап	III этап	
					Показатель травматизма на 100 работающих			Ожидаемое число	
	Число работающих	Число травм	Число работающих	Число травм	Первый период	Второй период	стандарт	Первый период	Второй период
Мужчины	200	32	600	72	16	12	400	64	48
Женщины	400	28	200	16	7	8	300	21	24
Всего	600	60	800	88	10	11	700	85	72
	IV этап. Определение стандартизованных показателей						100	12,1	10,3

V этап. Сопоставление соотношения интенсивных и стандартных показателей травматизма в периодах А и Б (табл. 6).

Показатели:	Первый период «А»	Второй период «Б»	Соотношение А и Б
Интенсивные	10,0	11,0	$A < B$
Стандартные	12,1	10,3	$A > B$

Выводы.

Анализ травматизма в первом и втором периодах времени выявил следующее:

- показатель травматизма на 100 работающих в целом выше во втором периоде (Б) чем в первом(А) ($11,0 > 10,0$);

- однако если бы гендерный состав работающих в промышленных предприятиях в первом и втором периодах был одинаков, то травматизм был бы выше во втором периоде.

- следовательно, на различия в уровнях травматизма (в частности, на "завышение" ее в группе «Б» и "занижение" в группе «А») оказала влияние неоднородность гендерного состава работающих, а именно, преобладание во втором периоде мужчин с относительно высоким показателем травматизма.

Стандартизованный показатель травматизма в первом периоде «Б» выше.

Таким образом, если бы составы больных в первом и втором периодах были одинаковыми, то травматизм был бы выше во втором периоде «Б».

Задача 3 – эталон.

На основе приведенных в табл. 1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.
2. Сравнить:
 - общие показатели, рассчитанные обычным способом;
 - частные показатели в отдельных группах;
 - стандартизованные и обычные показатели.
3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Вариант 3.

Таблица 1.

Распределение рабочих и длительно и часто болеющих (ДЧБ) лиц основных и вспомогательных цехов завода **Н.** по возрасту (в абс. числах).

Возраст в годах	Основные цеха (ОЦ)		Вспомогательные цеха (ВЦ)	
	Число рабочих	Из них ДЧБ	Число рабочих	Из них ДЧБ
20-29	100	5	200	21
30-39	300	45	150	25
40 и старше	200	40	50	10
Всего	600	90	400	56

За стандарт принять полусумму составов рабочих по возрасту в основных и вспомогательных цехах

Решение:

Этапы расчета стандартизованных показателей

I этап. Сначала определяют общие показатели ДЧБ лиц в двух сравниваемых совокупностях (в основных и вспомогательных цехах).

Основные цеха: $\frac{90 \times 100}{600} = 15$ на 100 работающих;

Вспомогательные цеха: $\frac{56 \times 100}{400} = 14$ на 100 работающих.

Затем находят показатели ДЧБ в зависимости от возраста рабочих.

Например: в основных цехах рабочие ДЧБ в возрасте 20-29 лет составляют $\frac{5 \times 100}{100} = 5$ на 100 работающих в возрасте 20-29 лет, а в вспомогательных цехах соответственно $= \frac{21 \times 100}{200} = 10,5$ (табл. 2).

Аналогично проводят расчеты и в других возрастных группах (см. сводную табл. 2 – I этап).

Таблица 2.

Показатели ДЧБ лиц в зависимости от возраста в основных и вспомогательных цехах:

Возраст в годах	Основные цеха (ОЦ)			Вспомогательные цеха (ВЦ)		
	Число рабочих	Из них ДЧБ	Показатель ДЧБ лиц в ОЦ на 100 работающих	Число рабочих	Из них ДЧБ	Показатель ДЧБ лиц в ВЦ на 100 работающих
20-29	100	5	5	200	21	10,5
30-39	300	45	15	150	25	16,6
40 и старше	200	40	20	50	10	20
Всего	600	90	15,0	400	56	14,0

II этап. Определение стандарта.

За стандарт принимают полусумму состава основных и вспомогательных цехов по возрасту (табл. 3).

Таблица 3.

Пол	Основные цеха (ОЦ)	Вспомогательные цеха (ВЦ)	Число работающих в основных и вспомогательных цехах	Стандарт
	Число работающих	Число работающих		
20-29	100	200	100+200	300
30-39	300	150	300+150	450
40 и старше	200	50	200+50	250
Всего	600	400	600+400	1000

III этап. Расчет ожидаемых чисел (в данном случае лиц ДЧБ) в каждой группе стандарта.

Если из 100 работающих ДЧБ 5, то, сколько бы составило число ДЧБ, если бы число работающих составляло 300 человек (стандарт).

Составляют пропорцию: 100 – 5

$$300 - X \quad X = \frac{5 \times 300}{100} = 15$$

Из данной пропорции получаем ожидаемую величину. Аналогично полученные данные в целом представлены в таблице 4.

Находят сумму ожидаемых чисел ДЧБ в основных (15 + 67,5 + 50 = 132,5) и вспомогательных цехах (31,5 + 75 + 2 = 108,5) (см. табл. 4).

Таблица 4.

Расчет ожидаемых величин ДЧБ в основных и вспомогательных цехах:

Возраст до 20-29 лет	Возраст от 30 до 39	Возраст 40 лет и старше
Основные цеха 100 – 5 300 – X $X = \frac{5 \times 300}{100} = 15$	Основные цеха 300 – 45 450 – X $X = \frac{45 \times 450}{300} = 67,5$	Основные цеха 200 – 40 250 – X $X = \frac{40 \times 250}{200} = 50$
Вспомогательные цеха 200 – 21 300 – X $X = \frac{21 \times 300}{200} = 31,5$	Вспомогательные цеха 150 – 25 450 – X $X = \frac{25 \times 450}{150} = 75$	Вспомогательные цеха 50 – 10 250 – X $X = \frac{10 \times 50}{250} = 2$

IV этап. Расчет стандартизованных коэффициентов.

Определяют общие стандартизованные показатели ДЧБ в основных и вспомогательных цехах.

При условии, что в каждом периоде число работающих составляло 1000 (стандарт, см. табл. 3), рассуждаем следующим образом: из 1000 работающих в ОЦ ожидаемое число ДЧБ составляет 132,5, в ВЦ 108,5, следовательно, показатель больных вычисляется на основе пропорции:

Основные цеха: $132,5 \times 100 / 1000 = 13,2$ на 100 работающих.

Вспомогательные цеха: $108,5 \times 100 / 1000 = 10,8$ на 100 работающих.

Это и есть стандартизованные показатели, т.е. показатели, вычисленные при условии, что состав работающих в ОЦ и ВЦ одинаковый (стандартный).

Результаты поэтапного расчета стандартизованных показателей лиц ДЧБ оформляют в виде таблицы (табл. 5).

Таблица 5.

Результаты поэтапного расчета стандартизованных показателей травматизма

Пол	Основные цеха (ОЦ)		Вспомогательные цеха (ВЦ)		I этап		II этап	III этап	
					Показатель ДЧБ на 100 работающих			Ожидаемое число	
	Число работающих	Из них ДЧБ	Число работающих	Из них ДЧБ	Основные цеха	Вспомогательные цеха	стандарт	Основные цеха	Вспомогательные цеха
20-29 лет	100	5	200	21	5	10,5	300	15	48
30-39 лет	300	45	150	25	15	16,6	450	67,5	24
40 и старше	200	40	50	10	20	20	250	50	72
Всего	600	90	400	56	15,0	14,0	1000	132,5	108,5
	IV этап. Определение стандартизованных показателей						100	13,2	10,8

V этап. Сопоставление соотношения интенсивных и стандартных показателей лиц ДЧБ в основных и вспомогательных цехах (табл. 6).

Таблица 6.

Показатели:	Основные цеха	Вспомогательные цеха	Соотношение ОЦ и ВЦ
Интенсивные	15,0	14,0	ОЦ > ВЦ
Стандартные	13,2	10,8	ОЦ > ВЦ

Выводы.

1. Уровень интенсивных показателей заболеваемости (ДЧБ) в целом в основных цехах выше, чем вспомогательных.
2. Однако уровень стандартных показателей лиц ДЧБ в целом в ОЦ и ВЦ ниже интенсивных.
3. Следовательно, анализ данных интенсивных и стандартизованных показателей заболеваемости (ДЧБ) подтверждает, что в основных цехах заболеваемость (ДЧБ) выше, чем – в вспомогательных.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Вычисление стандартизованных показателей

Задача 1.

На основе приведенных в табл.1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.
2. Сравнить:
 - общие показатели, рассчитанные обычным способом;
 - частные показатели в отдельных группах;
 - стандартизованные и обычные показатели.
3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Задание. Распределение населения городов А и Б по возрасту и числу умерших (в абс. числах) (табл. 1)

Таблица 1.

Возраст больных (в годах)	Город А		Город Б	
	число жителей	из них умерло	число жителей	из них умерло
0-14	3000	30	1000	10
15-49	5000	10	5000	10
50 и старше	2000	60	4000	120
Всего:	10000	10	10000	140

За стандарт принять полусумму состава жителей по возрасту в городах А и Б.

Задача 2.

На основе приведенных в табл.1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.
2. Сравнить:
 - общие показатели, рассчитанные обычным способом;
 - частные показатели в отдельных группах;
 - стандартизованные и обычные показатели.
3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Задание. Распределение городского и сельского населения по возрасту и числу умерших (в абс. числах) (табл. 1)

Таблица 1.

Возраст больных (в годах)	Городское население		Сельское население	
	Численность населения	из них умерло	Численность населения	из них умерло
0-14	50000	600	100000	1400
15-49	100000	400	400000	1600
50 и старше	50000	1000	300000	5400
Всего:	200000	2000	800000	8400

За стандарт принять полусумму состава городского и сельского населения по возрасту.

Задача 3.

На основе приведенных в табл.1 данных требуется:

3. Вычислить стандартизованные показатели, используя прямой метод стандартизации.

4. Сравнить:

- общие показатели, рассчитанные обычным способом;
- частные показатели в отдельных группах;
- стандартизованные и обычные показатели.

3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Задание. Распределение больных и умерших в двух больницах в зависимости от срока госпитализации при аппендиците (в абс. числах) (табл. 1).

Таблица 1.

Срок госпитализации в днях	Больница №1		Больница №2	
	Число больных	из них умерло	Число больных	из них умерло
1-3	400	1	100	-
3-4	550	3	200	2
5-6	50	6	300	10
Всего:	600	10	600	12

За стандарт принять сумму составов больных по срокам госпитализации в больницах №1 и №2.

Задача 4.

На основе приведенных в табл.1 данных требуется:

1. Вычислить стандартизованные показатели, используя прямой метод стандартизации.

2. Сравнить:

- общие показатели, рассчитанные обычным способом;
- частные показатели в отдельных группах;
- стандартизованные и обычные показатели.

3. Сделать вывод, вытекающий из сопоставления обычных и стандартизованных показателей в сравниваемых группах.

Задание. Распределение рабочих и лиц, получивших инвалидность от сердечно-сосудистых заболеваний, на двух предприятиях по полу (в абс. числах) (табл. 1)

Таблица 1.

	Предприятие №1		Предприятие №2	
	Число рабочих	Число инвалидов	Число рабочих	Число инвалидов
	400	10	200	7
	200	9	500	18
Всего:	600	19	700	25

За стандарт принять состав рабочих по полу (сумма по двум предприятиям).

Литература

Основная литература:

1. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения. Под ред. чл.-корр. РАМН, проф. В.З.Кучеренко. М., "Гэотар-Медиа", 2007, учебное пособие для вузов
2. Кобринский Б.А., Зарубина Т.В. Медицинская информатика: Учебник. М: изд. "Академия", 2009.
3. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: Учебное пособие для практических занятий /Под ред. В.З.Кучеренко. – М.:ГЭОТАР-МЕД, 2004. – 192 с.

Дополнительная литература:

4. Лисицын Ю.П. Общественное здоровье и здравоохранение. Учебник для вузов. — М.: ГЭОТАР-МЕД, 2007. — 512 с.
5. Медик В.А., Юрьев В.К. Курс лекций по общественному здоровью и здравоохранению: Часть 1. Общественное здоровье. — М.: Медицина, 2003. — 368 с.
6. Кучеренко В.З., Агарков Н.М. и др. Социальная гигиена и организация здравоохранения (Учебное пособие) — Москва, 2000. — 432 с.

