

**МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**  
**УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ**  
**«ГОМЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»**

**Кафедра общественного здоровья и здравоохранения**

**Л. Г. Соболева, В. М. Дорофеев, Т. М. Шаршакова**

**ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ  
ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.  
МЕТОД СТАНДАРТИЗАЦИИ. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ**

**Учебно-методическое пособие для студентов 4 курса  
высших медицинских учебных заведений всех факультетов**

**Гомель  
ГоГМУ  
2009**

УДК 614.2:31  
ББК 51.1(2)+60.5  
С 54

Рецензент:

кандидат медицинских наук, доцент, кафедры общей гигиены, экологии и радиационной медицины Гомельского государственного медицинского университета *С. М. Дорофеева*

**Соболева, Л. Г.**

**С 54** Параметрические методы оценки достоверности статистического исследования. Метод стандартизации. Корреляционный анализ: учеб.-метод. пособие для студентов 4 курса высших медицинских учебных заведений всех факультетов / Л. Г. Соболева, В. М. Дорофеев, Т. М. Шаршаква. — Гомель: Учреждение образования «Гомельский государственный медицинский университет», 2009. — 24 с.

ISBN 978-985-506-217-3

Предназначено для проведения практических занятий на кафедре общественного здоровья и здравоохранения с целью получения знаний, умений и навыков по данной теме. Соответствует учебному плану и типовой учебной программе по дисциплине «Общественное здоровье и здравоохранение», утвержденной Министерством здравоохранения Республики Беларусь.

Утверждено и рекомендовано к изданию Центральным учебным научно-методическим советом учреждения образования «Гомельский государственный медицинский университет» 19 декабря 2008 г., протокол № 12.

**УДК 614.2:31  
ББК 51.1(2)+60.5**

**ISBN 978-985-506-217-3**

© Учреждение образования  
«Гомельский государственный  
медицинский университет», 2009

## ВВЕДЕНИЕ

Тема изучается в течение 5 часов и состоит из 3 частей.

В первой части занятия разбираются теоретические аспекты данной темы, рассматриваются методики правильного расчета коэффициента корреляции, стандартизованных коэффициентов, методы оценки достоверности результатов статистического исследования. Вторая часть посвящена самостоятельной работе студентов. Студенты решают ситуационные задачи по оценке достоверности результатов статистического исследования, корреляционному анализу, вычислению стандартизованных показателей. В третьей части занятия проводится оценка полученных знаний путем проведения компьютерного тестирования.

## ЦЕЛЬ ЗАНЯТИЯ

Обучить студентов основам корреляционного анализа, расчета стандартизованных коэффициентов. При изучении этой темы студенты должны усвоить методы оценки достоверности результатов статистического исследования.

## ЗАДАЧИ ЗАНЯТИЯ:

- усвоить сущность корреляционной связи между признаками;
- овладеть методикой вычисления коэффициента корреляции по методу квадратов Пирсона и оценки характера, силы и достоверности связи;
- усвоить основные положения метода стандартизации статистических показателей;
- научиться вычислять стандартизированные показатели;
- научиться оценивать достоверность результатов статистического исследования.

## ТРЕБОВАНИЯ К ИСХОДНОМУ УРОВНЮ ЗНАНИЙ

**Студент должен уметь:**

- вычислять коэффициент корреляции;
- оценивать достоверность коэффициента корреляции;
- вычислять стандартизованные показатели прямым методом.

**Студент должен знать:**

- критерии достоверности;
- доверительные границы средних и относительных величин;
- достоверность разности между двумя средними величинами;
- виды связи между явлениями и признаками;
- дать определение понятия «корреляционная связь»;
- оценка достоверности коэффициента корреляции;
- сущность стандартизации показателей;
- виды метода стандартизации.

## КОНТРОЛЬНЫЕ ВОПРОСЫ ИЗ СМЕЖНЫХ ДИСЦИПЛИН

1. Как в математике и физике выражается связь между признаками?
2. Что такое полная, или функциональная связь между признаками?

## КОНТРОЛЬНЫЕ ВОПРОСЫ ПО ТЕМЕ ЗАНЯТИЯ

1. Оценка степени достоверности результатов статистического исследования.
2. Виды связи между явлениями или признаками.
3. Корреляционная связь.
4. Критерии оценки характера и силы корреляции.
5. Методы определения коэффициента корреляции.
6. Применение метода стандартизации.
7. Методы стандартизации.
8. Этапы прямого метода стандартизации.
9. Стандартизованные показатели.

## УЧЕБНЫЙ МАТЕРИАЛ

### ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Задачей статистического исследования является выявление закономерностей, лежащих в природе исследуемых явлений. Показатели и средние величины должны служить отображением действительности, для чего необходимо определять степень их достоверности. Правильное отображение выборочной совокупностью генеральной совокупности называется репрезентативностью. Мерой точности и достоверности выборочных статистических величин являются средние ошибки представительности (репрезентативности), которые зависят от численности выборки и степени разнообразия выборочной совокупности по исследуемому признаку.

Поэтому для определения степени достоверности результатов статистического исследования необходимо для каждой относительной и средней величины вычислить соответствующую среднюю ошибку. Средняя ошибка показателя  $m_p$  вычисляется по формуле:

$$m_p = \pm \sqrt{\frac{P \cdot q}{n}}.$$

При числе наблюдений менее 30  $m_p = \pm \sqrt{\frac{P \cdot q}{n-1}}$ ,

где  $P$  — величина показателя в процентах, промилле и т. д.;

$q$  — дополнение этого показателя до 100, если он в процентах, до 1000, если ‰ и т.д. (т. е.  $q = 100 - P$ ,  $1000 - P$  и т. д.).

Например, известно, что в районе в течение года заболело дизентерией 224 человека. Численность населения — 33000. Показатель заболеваемости дизентерией на 10000 =  $\frac{244 \cdot 10000}{33000} = 74$ .

$$\text{Средняя ошибка этого показателя } m_p = \pm \sqrt{\frac{74 \cdot (10000 - 74)}{33000}} = \pm 4,79.$$

Для решения вопроса о степени достоверности показателя определяют доверительный коэффициент (t), который равен отношению показателя к его средней ошибке, т. е.

$$t = \frac{P}{m_p}.$$

$$\text{В нашем примере } t = \frac{74}{4,79} = 15,4.$$

Чем выше t, тем больше степень достоверности. При t=1, вероятность достоверности показателя равна 68,3 %, при t=2 — 95,5 %, при t=3 — 99,7 %. В медико-статистических исследованиях обычно используют доверительную вероятность (надежность), равную 95,5–99,0 %, а в наиболее ответственных случаях — 99,7 %. Таким образом, в нашем примере показатель заболеваемости достоверен.

При числе наблюдений менее 30, значение критерия определяется по таблице Стьюдента. Если полученная величина будет выше или равна табличной — показатель достоверен. Если ниже — не достоверен.

При необходимости сравнения двух однородных показателей достоверность их различий определяется по формуле:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \text{ (от большего числа отнимают меньшее),}$$

где  $P_1 - P_2$  — разность двух сравниваемых показателей,

$\sqrt{m_1^2 + m_2^2}$  — средняя ошибка разности двух показателей.

Например, в районе Б в течении года заболело дизентерией 270 человек. Население района — 45000. Отсюда заболеваемость дизентерией:

$$P = \frac{270 \cdot 10000}{45000} = 60.$$

$$m_p = \pm \sqrt{\frac{60 \cdot (10000 - 60)}{45000}} = \pm 3,64;$$

$$t = \frac{60}{3,64} = 16,48 \text{ т. е. показатель заболеваемости достоверен.}$$

Как видно, заболеваемость в районе Б ниже, чем в районе А. Определяем по формуле достоверность разницы двух показателей:

$$t = \frac{74 - 60}{\sqrt{4,79^2 + 3,64^2}} = \frac{14}{6} = 2,3.$$

При наличии большого числа наблюдений (более 30) разность показателей является статистически достоверной, если  $t = 2$  или больше. Таким образом, в нашем примере заболеваемость в районе А достоверно выше, т. к. доверительный коэффициент ( $t$ ) больше 2.

Зная величину средней ошибки показателя, можно определить доверительные границы этого показателя в зависимости от влияния причин случайного характера. Доверительные границы определяются по формуле:

$$P \pm t \cdot m,$$

где  $P$  — показатель;

$m$  — его средняя ошибка;

$t$  — доверительный коэффициент выбирается в зависимости от требуемой величины надежности:  $t = 1$  соответствует надежности результата в 68,3 % случаев,  $t = 2$ —95,5 %,  $t = 2,6$ —99 %,  $t = 3$ —99,7 %,  $t = 3,3$ —99,9. Величина  $t \cdot m$  называется предельной ошибкой.

Например, в районе Б показатель заболеваемости дизентерией с точностью до 99,79 % может колебаться в связи со случайными факторами в пределах  $60 \pm 3 \cdot 3,64$  т. е. от 49,1 до 70,9.

Также, как и для относительных величин необходимо для средней арифметической определять ее среднюю ошибку.

Средняя ошибка средней арифметической  $m_x$  определяется по формуле:

$$m_x = \pm \frac{\sigma}{\sqrt{n}}$$

или

$$m_x = \pm \frac{\sigma}{\sqrt{n-1}} \text{ (при числе наблюдений менее 30),}$$

где  $\sigma$  — среднее квадратичное отклонение;

$n$  — число наблюдений.

Определение средней ошибки средней арифметической необходимо:

1. Для оценки достоверности средней арифметической, которая определяется по формуле:

$$t = \frac{M}{m}.$$

2. Для определения достоверности разности двух средних арифметических, которая определяется по формуле:

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

В обоих случаях оценка производится так же, как и при определении достоверности показателей, т.е. при числе наблюдений более 30 результаты достоверны, если величина доверительного коэффициента (t) равна или больше 2.

3. Для определения доверительных границ средней арифметической, т.е. в каких пределах может колебаться средняя арифметическая в зависимости от случайных факторов.

Например, обследовано 15 рабочих в цехе с высокой температурой окружающего воздуха на частоту пульса. Результаты отражены в таблица 1.

Таблица 1 — Результаты обследования

Число ударов пульса X V	Число рабочих (P)	VP	V-M	(V-M) <sup>2</sup>	(V-M) <sup>2</sup> P
76	3	228	-5,3	28,09	84,3
78	5	390	-3,3	10,89	55,0
82	3	246	0,7	0,49	1,5
88	2	176	6,7	44,89	90,0
90	2	180	8,7	75,69	151,4
15	1220382,2				

**Определить среднюю частоту пульса и ее достоверность.**

Вначале по формуле определяем среднюю арифметическую:

$$M = \frac{\sum VP}{n} = \frac{1220}{15} = 81,3 \text{ уд. в мин.}$$

По формуле определяем среднее квадратичное отклонение:

$$\sigma = \pm \sqrt{\frac{\sum (V - M)^2 P}{\sum P}} = \pm \sqrt{\frac{382,2}{15 - 1}} = \pm \sqrt{27,3} = \pm 5,2.$$

Затем по формуле вычисляем среднюю ошибку средней арифметической

$$m = \pm \frac{\sigma}{\sqrt{n - 1}} = \pm \frac{5,2}{\sqrt{14}} = \pm \frac{5,2}{3,74} = \pm 1,4.$$

Для оценки степени достоверности средней арифметической полученные данные подставляем в формулу:

$$t = \frac{M}{m} = \frac{81,3}{1,4} = 58,1.$$

Величина достоверна.

Допустим известно, что в цехе с нормальной температурой окружающего воздуха (18–20°) при обследовании 25 рабочих установлена средняя частота пульса 72,4 уд. в мин. Среднее квадратичное отклонение — 4,8 уд. в мин, средняя ошибка 0,96. Требуется определить достоверность разности в частоте пульса у рабочих обоих цехов. Для этого данные подставляем в формулу:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{81,3 - 72,4}{\sqrt{1,4^2 + 0,96^2}} = \frac{8,9}{\sqrt{1,96 + 0,9216}} = \frac{8,9}{1,7} = 5,2.$$

Разница в частоте ударов пульса у рабочих сравниваемых цехов статистически достоверна.

Наконец, определяем доверительные границы частоты пульса у рабочих в цехе с высокой температурой окружающего воздуха со степенью вероятности 99,7 %.  $M \pm 3m$  т. е. от 77,1 до 85,5 уд. в мин.

## ИЗМЕРЕНИЕ СВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ

Все явления в природе и обществе находятся во взаимной связи. По характеру зависимости явлений различают:

- функциональную (полную);
- корреляционную (неполную) связи.

*Функциональная связь* означает строгую зависимость явлений, когда любому значению одного из них всегда соответствует определенное одно и то же значение другого. Функциональная связь обычно выражается формулами:

$$\text{Объем тела} = \frac{\text{масса}(m)}{\text{плотность}(p)}; \quad \text{путь} = \text{скорость} \cdot \text{время} \text{ и т. д.}$$

*При корреляционной же связи* одной и той же величине одного признака соответствуют разные величины другого. Например, между ростом и весом имеется корреляционная связь, между заболеваемостью злокачественными новообразованиями и возрастом и т. д.

По направлению различают прямые и обратные корреляционные связи. При прямой — увеличение одного из признаков ведет к увеличению другого; при обратном же — с увеличением одного признака второй уменьшается.

По силе связь может быть сильной, средней и слабой.

На основе статистического анализа можно установить наличие связи, ее направление и измерить ее силу. Одним из способов измерения связи между явлениями является вычисление коэффициента корреляции, который обозначается  $r_{xy}$ . Он может быть вычислен различными способами. Наиболее точным является метод квадратов (Пирсона), при котором коэффициент корреляции определяется по формуле:

$$r_{xy} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}},$$

где  $r_{xy}$  — коэффициент корреляции между статистическим рядом X и Y;  
 $d_x$  — отклонение каждого из чисел статистического ряда X от своей средней арифметической;

$d_y$  — отклонение каждого из чисел статистического ряда Y от своей средней арифметической.

В зависимости от силы связи и ее направления коэффициент корреляции может находиться в пределах от 0 до 1 (-1). Коэффициент корреляции, равный 0, говорит о полном отсутствии связи. Чем ближе уровень коэффициента корреляции к 1 или (-1), тем соответственно больше, теснее измеряемая им прямая или обратная связь. При коэффициенте корреляции равном 1 или (-1) связь полная, функциональная (таблица 2).

Таблица 2 — Схема оценки силы корреляционной связи по коэффициенту корреляции

Сила связи	Величина коэффициента корреляции при наличии	
	прямой связи (+)	обратной связи (-)
Связь отсутствует	0	0
Связь малая (слабая)	от 0 до +0,29	от 0 до -0,29
Связь средняя (умеренная)	от +0,3 до +0,69	от -0,3 до -0,69
Связь большая (сильная)	от +0,7 до +0,99	от -0,7 до -0,99
Связь полная (функциональная)	+1,0	-1,0

Для вычисления коэффициента корреляции по методу квадратов составляется таблица из 7 колонок. Разберем процесс вычисления на примере.

Определить силу и характер связи между содержанием йода в воде и пораженностью зубом (таблица 3).

Таблица 3 — Исходные данные.

Содержание йода в мг в воде ( $V_x$ )	Пораженность зубом в % ( $V_y$ )	$d_x = V_x - M_x$	$d_y = V_y - M_y$	$d_x d_y$	$d_x^2$	$d_y^2$
201	0,2	63	-3,6	-226,8	3969	0,04
178	0,6	40	-3,2	-128,0	1600	0,36
155	1,1	17	-2,7	-45,9	289	1,21
154	0,8	16	-3,0	-48,0	256	0,64
126	2,5	-12	-1,3	-15,6	144	6,25
81	4,4	-57	0,6	-34,2	3249	19,36
71	16,9	-67	13,1	-877,7	4489	285,61
				<b><math>\Sigma -1345,0</math></b>	<b><math>\Sigma 13996,0</math></b>	<b><math>\Sigma 313,47</math></b>

1. Определяем среднее содержание йода в воде (в мг/л):

$$M_x = \frac{\sum V_x}{n} = \frac{996}{7} = 138 \text{ мг/л.}$$

2. Определяем среднюю пораженность зубом в %:

$$M_y = \frac{\sum V_y}{n} = \frac{26,5}{7} = 3,8\%.$$

3. Определяем отклонение каждого  $V_x$  от  $M_x$ , т.е.  $d_x$ .

201-138=63; 178-138=40 и т.д.

4. Аналогично определяем отклонение каждого  $V_y$  от  $M_y$ , т. е.  $d_y$ .  
 $0,2-3,8 = -3,6$ ;  $0,6-3,8 = -3,2$  и т. д.

5. Определяем произведения отклонений. Полученное произведение суммируем и получаем:

$$\sum d_x d_y = -1345,0$$

6.  $d_x$  возводим в квадрат и результаты суммируем, получаем:

$$\sum d_x^2 = 13966,0$$

7. Аналогично возводим в квадрат  $d_y$ , результаты суммируем, получим:

$$\sum d_y^2 = 313,47$$

8. Наконец, все полученные суммы подставляем в формулу:

$$r_{xy} = \frac{-1345}{\sqrt{13966 \cdot 313,47}} = \frac{-1345}{2094,59} = -0,64$$

Для решения вопроса о достоверности коэффициента корреляции определяют его среднюю ошибку по формуле:

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

(Если число наблюдений менее 30, тогда в знаменателе  $n-1$ ).

В нашем примере

$$m_r = \frac{1 - 0,64^2}{\sqrt{7-1}} = \frac{1 - 0,4096}{2,45} = 0,24.$$

Величина коэффициента корреляции считается достоверной, если не менее чем в 3 раза превышает свою среднюю ошибку.

В нашем примере  $\frac{0,64}{0,24} = 2,7$ .

Таким образом, коэффициент корреляции не достоверен, что вызывает необходимость увеличения числа наблюдений.

Коэффициент корреляции можно определить несколько менее точным, но намного более легким способом — методом рангов (Спирмена).

## СТАНДАРТИЗОВАННЫЕ ПОКАЗАТЕЛИ

При анализе материалов о здоровье населения, деятельности лечебно-профилактических организаций, качестве оказания медицинской помощи, эффективности лечебных и оздоровительных мероприятий применяется сравнение общих интенсивных показателей или средних величин. Но это сравнение дает истинный результат лишь в тех случаях, если показатели вычислены в отношении качественно однородной совокупности (по возрасту, полу, профессии и т. д.), поскольку на величину общих интенсивных показателей влияет не только разная санитарная обстановка или качество лечебно-профилактической работы, но и различие в возрастном, половом, профессиональном, социальном составе сравниваемых групп населения.

Наибольшее влияние на уровень многих показателей оказывает возрастной состав населения. Так, известно, что наиболее высокие показатели заболеваемости и смертности отмечаются у детей и стариков. Ниже у лиц молодого и среднего возраста. Следовательно, там, где в составе населения больше детей и стариков, уровень заболеваемости и смертности бывает выше, даже если санитарные условия и медицинская помощь лучше (таблица 4).

Таблица 4 — Летальность от очаговой пневмонии в двух детских больницах

Возраст детей, лет	Больница № 1			Больница № 2		
	число больных	число умерших	летальность	число больных	число умерших	летальность
0–3	1500	90	6,0	500	40	8,0
4–7	500	10	2,0	500	15	3,0
Старше	500	5	1,0	1500	22	1,5
ВСЕГО	2500	105	4,2	2500	77	3,1

Как видно из таблицы, в больнице № 1 общий показатель летальности выше, чем в больнице № 2, несмотря на то, что повозрастные показатели ниже.

Объясняется это тем, что состав больных по возрасту в сравниваемых больницах различен: в больнице № 1 детей в возрасте до 3 лет лечилось 60 %, а в больнице № 2 только 20 %, а дети младшего возраста имеют более высокие показатели летальности. Поэтому на основании общих показателей летальности можно делать выводы при сравнении двух больниц лишь при условии однородного состава отделений, примерно одинакового состава больных и т.д. В тех же случаях, когда приходится сравнивать интенсивные показатели в двух совокупностях, состав которых не одинаков, правильные выводы могут быть получены только на основании специально вычисляемых показателей.

Статистический метод, позволяющий исключить (элиминировать) влияние неоднородного состава сравниваемых групп на общие показатели, называется методом стандартизации. В результате использования этого метода получают стандартизованные показатели, т. е. условные, которые могли бы быть при условии однородного состава. **Показатели эти применяются только в целях сравнения.** При изменении стандарта меняется и величина стандартизованных показателей, поэтому при анализе важны не их абсолютные величины, а степень различия стандартизованных показателей между собой.

В зависимости от наличия исходных данных можно пользоваться одним из трех методов стандартизации: прямым, косвенным или обратным.

**Прямой метод стандартизации** — наиболее распространенный.

Стандартизованные показатели вычисляются прямым методом, когда известны:

- состав совокупности (населения, работающих, больных);
- состав изучаемого явления по этим же группам.

Стандартизация проводится в следующей последовательности:

1. Вычисление групповых показателей.
2. Выбор и вычисление стандарта.

В качестве стандарта можно взять:

- состав одной из сравниваемых групп;
- средний состав или состав обеих групп, взятых вместе;
- состав третьего объекта, известного по предыдущим исследованиям или по другим материалам.

Желательно состав третьего объекта брать тот, в состав которого входят сравниваемые группы.

Например, при сравнении заболеваемости или смертности в двух районах за стандарт целесообразно взять возрастной состав населения области, республики.

3. Вычисление «ожидаемых» чисел изучаемого явления в сравниваемых совокупностях по стандарту.

4. Определение стандартизованных показателей и выводы.

Сущность прямого метода состоит в том, что условно принимают какой-либо состав населения или больных за стандарт и считают его одинаковым в сравниваемых совокупностях. Затем, учитывая действительные размеры явления по групповым показателям, вычисляют стандартизованные коэффициенты.

Прямой метод стандартизации состоит из следующих этапов:

*I этап* — вычисление интенсивных показателей (таблица 5).

В нашем примере — вычисление частоты случаев нетрудоспособности в целом по цехам и в каждой возрастной группе.

Рассчитываем общий интенсивный показатель для цеха № 1:

1000 работающих — 1345 случаев,

100 работающих —  $x$ ,

$$x = \frac{1345 \times 100}{1000} = 134,5.$$

Таблица 5 — I этап прямого метода стандартизации

Возрастные группы (в годах)	Цех № 1		Цех № 2		Частота сл. нетруд. на 100 работающих		Стандарт. (число работающих)	«Ожидаемое» число случаев нетрудоспособности	
	число работающих	число случаев нетрудоспособности	число работающих	число случаев нетрудоспособности	Цех № 1	Цех № 2		Цех № 1	Цех № 2
до 19	120	80	200	140	66,7	70,0	320	213,44	224,0
20–39	380	400	500	575	105,3	115,0	880	926,64	1012,0
40–59	400	680	80	132	170,0	165,0	480	816,00	792,0
60 и старше	100	185	20	34	185,0	170,0	120	222,00	204,0
ИТОГО:	1000	1345	800	881	134,5	110,1	1800	2178,0	2232,0

В возрасте до 19 лет (для цеха № 1) — частный интенсивный показатель:  
 120 работающих — 80 случаев;  
 100 работающих —  $x$ ;  
 $x = \frac{80 \times 100}{120} = 66,7$  и т.д. (для цеха № 1 и цеха № 2).

*II этап* прямого способа стандартизации — выбор стандарта. За стандарт обычно берут распределение по устраняемому фактору одной из сред, либо их сумму, либо подсумму. Однако, стандарт можно выбрать любой, по нашему смотрению. В нашем примере за стандарт можно взять распределение работающих в цехе № 1, либо в цехе № 2, либо средний возрастной состав по цехам № 1 и № 2, либо любое другое распределение работающих по возрасту. Пусть за стандарт мы возьмем сумму работающих в обоих цехах, т.е. в возрасте до 19 лет:  $120 + 200 = 320$  работающих, в 20–39 лет —  $380 + 500 = 880$  работающих и т. д.

*III этап* — наиболее ответственный — вычисление ожидаемых чисел. Необходимо вычислить, какова была бы величина сравниваемого явления, если бы частные интенсивные показатели остались прежними, а распределение среды было бы таким, как по стандарту.

В нашем примере вычисляем, сколько было бы случаев нетрудоспособности в цехе № 1 и № 2, если бы повозрастные показатели работающих в каждой возрастной группе было бы таким, как по стандарту и одинаковыми (что очень важно) в цехе № 1 и № 2. Найдем, сколько случаев нетрудоспособности было бы у работающих в возрасте до 19 лет в цехе № 1:

на 100 работающих — 66,7 случаев;

на 320 работающих —  $x$  случаев;

$$x = \frac{66,7 \times 320}{100} = 213,44.$$

В том же цехе № 1 в возрасте 20–39 лет число случаев нетрудоспособности составило бы:

на 100 работающих — 105,3 случая;

на 880 работающих —  $x$  случаев;

$$x = \frac{105,3 \times 880}{100} = 926,64 \text{ и т. д.}$$

*IV этап* — вычисление стандартизованных показателей и их сравнение.

Стандартизованные показатели рассчитываются, как и обычные интенсивные показатели, на величину стандарта. В нашем примере рассчитываем частоту случаев нетрудоспособности на 100 рабочих. В цехе № 1 на 1800 работающих было бы всего 2178,1 случаев нетрудоспособности, следовательно, на 100 работающих —  $x$ .

$$x = \frac{2178,1 \times 100}{1800} = 121,0.$$

В цехе № 2:

$$x = \frac{2232 \times 100}{1800} = 124,0.$$

**Выводы.** Если бы возрастной состав работающих в цехах был одинаковым, то частота случаев заболеваемости в цехе № 2 была бы выше, чем в цехе № 1. Более высокий интенсивный показатель частоты случаев нетрудоспособности в цехе № 1 (134,5) обусловлен тем, что в нем среди работающих лиц старше 40 лет составляют 50 %, а в цехе № 2 — только 12,5 %.

Стандартизованные показатели *могут быть использованы* только для сравнения, их величина является условной, зависящей от выбора условий (стандарта), она не дает представления об истинном размере того или иного явления. Поэтому указывать величину стандартизованного показателя не следует. На основании сравнения стандартизованных показателей можно судить, где выше или ниже величина явления при условии устранения влияния на него неоднородного состава сравниваемых совокупностей или (проще), как соотносились бы интенсивные показатели при устранении влияния на них определенного фактора.

**Косвенный метод** применяется при отсутствии данных о распределении изучаемого явления или очень малых числах при этом распределении. Для проведения стандартизации по этому методу нужны следующие данные:

- распределение совокупности по составу (населения по возрасту рабочих, по стажу работы и т. д.);
- общая численность изучаемого явления (число заболевших, умерших и т. п.);
- погрупповые интенсивные показатели, которые можно принять за стандарт (по области, республике и др).

**Обратный метод** применяется при отсутствии сведений о составе сравниваемых совокупностей. Для проведения стандартизации этим методом требуются данные:

- состав изучаемого в обоих совокупностях явления (больных, умерших и т. п.);
- общая численность сравниваемых совокупностей (населения, рабочих т. п.).
- погрупповые интенсивные показатели, которые можно взять за стандарт (по области, республике и др).

**Последовательность вычисления косвенного и обратного методов:**

- выбор стандарта;
- вычисление «ожидаемых» численностей обеих совокупностей (населения, больных);
- вычисление стандартизованных показателей.

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ ПО РЕАЛИЗАЦИИ ПРОГРАММНЫХ ЦЕЛЕЙ ЗАНЯТИЯ СИТУАЦИОННЫЕ ЗАДАЧИ

1. При изучении средней длительности лечения больных в хирургическом и терапевтическом отделениях городской больницы были получены следующие данные: средняя длительность лечения в хирургическом отделении составила 12,1 ( $m_1 = \pm 0,05$ ) дней, а в терапевтическом — 23,2 ( $m_2 = \pm 0,09$ ) дней. Общее число лечившихся составило — 405 человек.

Определить достоверность разности средней длительности лечения в сравниваемых группах.

3. Определить корреляционную связь между возрастом и числом госпитализированных больных с сердечной недостаточностью (таблица 6).

Таблица 6 — Исходные данные

Возраст в годах (X)	Число госпитализированных (Y)
До 20 лет	12
20–29	24
30–39	30
40–49	23
50–59	26
60 и старше	30

3. Дать сравнительный анализ деятельности по городским больницам № 1 и № 2, применив прямой метод стандартизации (таблица 7).

За стандарт принять число лечившихся в больнице № 1.

Таблица 7 — Исходные данные

Отделение	Больница № 1			Больница № 2		
	число лечив.	число умерших	летальность	число лечив.	число умерших	летальность
Терапевтическое	800	20	2,85	650	13	2,0
Хирургическое	700	11	1,6	450	5	1,1
Туберкулезное	500	30	6,0	900	45	5,0
Всего	2000	61	3,1	2000	63	3,2

4. Дать сравнительный анализ деятельности по двум городским больницам А и Б, применив метод стандартизации (таблица 8).

За стандарт принять структуру лечившихся в больнице А.

Таблица 8 — Исходные данные

Название болезни	Больница А			Больница Б		
	число больн.	число умерших	летальность %	число больн.	число умерших	летальность %
Гипертонич. болезнь	180	4	2,2	200	4	2,0
Рак желудка	100	30	30,0	90	27	30,0
Инфаркт миокарда	120	8	6,7	160	10	6,3
Всего	400	42	10,5	450	41	9,1

## САМОКОНТРОЛЬ УСВОЕНИЯ ТЕМЫ

Контрольные вопросы темы

1. Какие параметрические методы оценки достоверности Вы знаете?
2. Какие виды связи существуют между явлениями или признаками?
3. Что такое корреляционная связь (корреляция)?
4. Что является критерием оценки характера и силы корреляции?
5. Каковы методы определения коэффициента корреляции?
6. Как определяются характер и сила связи по коэффициенту корреляции?
7. Как оценивается достоверность коэффициента корреляции?
8. Для чего применяется метод стандартизации?
9. Какие существуют методы стандартизации?
10. В чем суть прямого метода стандартизации?
11. Когда применяют прямой метод стандартизации?
12. Назовите этапы прямого метода стандартизации.
13. Каковы способы выбора (или расчета) стандарта?
14. О чем свидетельствуют стандартизованные показатели?
15. Назовите случаи применения стандартизованных показателей в практической деятельности врача.

## ТЕСТЫ

**1. Величина стандартизованных показателей истинному размеру явлений:**

*Варианты ответа:*

- а) соответствует;
- б) не соответствует.

**2. Основное условие применения метода стандартизации:**

*Варианты ответа:*

- а) достаточное число наблюдений;
- б) наличие данных о структуре среды и явления;
- в) резкие различия состава изучаемых совокупностей.

**3. Стандартизованные показатели применяются:**

*Варианты ответа:*

- а) для характеристики первичного материала;
- б) для анализа и полученных данных;
- в) для сравнения между собой.

**4. Метод стандартизации применяется:**

*Варианты ответа:*

- а) для определения характера и силы связи между двумя явлениями (признаками);

б) при сравнении интенсивных показателей для устранения влияния на их величину определенных, факторов (признаков), обуславливавших качественную неоднородность совокупностей;

в) для определения достоверности различия двух сравниваемых показателей.

**5. Прямой метод стандартизации при сравнении показателей общей смертности населения двух городов можно применять, если:**

*Варианты ответа:*

а) известны состав населения по возрасту и состав умерших по возрасту в каждом из городов;

б) есть данные о распределении населения, имеются сведения об общей численности населения и о распределении умерших по возрасту в каждом из городов;

в) имеются данные о распределении населения по возрасту, общем числе умерших в каждом из городов, но сведений о распределении умерших по возрасту нет (или их число в каждой возрастной группе мало).

**6. Из применяемых и обозначенных ниже методов стандартизации наиболее точным является:**

*Варианты ответа:*

а) обратный метод;

б) косвенный метод;

в) прямой метод.

**7. При отсутствии данных о возрастном составе населения, когда имеются лишь сведения о возрастном составе больных или умерших, применяется метод стандартизации:**

*Варианты ответа:*

а) прямой;

б) обратный;

в) косвенный.

**8. При известном возрастном составе населения, а также повозрастной смертности от злокачественных новообразований применяется метод стандартизации:**

*Варианты ответа:*

а) прямой;

б) косвенный;

в) обратный.

**9. При сравнении показателей заболеваемости студентов двух вузов, если имеются данные распределения студентов по полу, но нет данных о распределении болевших по полу, следует применять метод стандартизации:**

*Варианты ответа:*

а) прямой;

- б) косвенный;
- в) обратный.

**10. Сравнить стандартизованные интенсивные показатели, если они исчислены с применением неодинакового стандарта:**

*Варианты ответа:*

- а) можно;
- б) нельзя;
- в) можно, но только при малом числе наблюдений;
- г) можно, но только при большом числе наблюдений.

**11. При сравнении интенсивных показателей, полученных на однородных по своему составу совокупностях, необходимо применять:**

*Варианты ответа:*

- а) оценку показателей соотношения;
- б) определение относительной величины;
- в) стандартизацию;
- г) оценку достоверности разности показателей.

**12. При вычислении стандартизованных показателей за стандарт можно применять:**

*Варианты ответа:*

- а) распределение одной из сравниваемых совокупностей;
- б) данные из других источников (отчеты, опубликованные материалы и др.);
- в) средний состав сравниваемых совокупностей;
- г) интенсивный показатель, характеризующий частоту явления (признака) в одной из сравниваемых совокупностей.

**13. Метод стандартизации применяется:**

*Варианты ответа:*

- а) при сравнении показателей заболеваемости гипертонической болезнью рабочих двух однотипных предприятий с резко различающимся половым составом рабочих;
- б) при сравнении показателей заболеваемости населения с разным возрастным составом в трех городах;
- в) при сравнении общих показателей летальности в двух больницах, в которых имеются профильные отделения (терапевтическое, хирургическое, инфекционное) и различное распределение больных по отделениям;
- г) при сравнении структуры причин младенческой смертности за разные годы;
- д) при изучении эффективности того или иного метода лечения с использованием неоднородных по возрасту основной и контрольных групп больных.

**14. Сколько существует методов стандартизации?**

*Варианты ответа:*

- а) 3;

- б) 2;
- в) 1.

**15. Стандартизованные коэффициенты являются:**

*Варианты ответа:*

- а) условными;
- б) средними.

**16. Стандартизованные коэффициенты применяется исключительно в целях:**

*Варианты ответа:*

- а) сравнения;
- б) изучения.

**17. Чаще всего используется наиболее простой метод стандартизации:**

*Варианты ответа:*

- а) прямой;
- б) косвенный;
- в) обратный.

**18. Сколько существует этапов вычисления стандартизованных показателей прямым методом?**

*Варианты ответа:*

- а) 3;
- б) 4;
- в) 2.

**19. На втором этапе стандартизации при прямом методе выбирается:**

*Варианты ответа:*

- а) стандарт;
- б) средняя арифметическая.

**20. Первый этап при косвенном методе стандартизации состоит в выборе:**

*Варианты ответа:*

- а) стандарта;
- б) статистического показателя.

**21. Стандартизованные коэффициенты применяются:**

*Варианты ответа:*

- а) для определения достоверности различия сравниваемых показателей;
- б) для определения характера и силы связи между явлениями;
- в) для сравнения интенсивных показателей, полученных на неоднородных по возрасту совокупностях.

**22. Функциональной связью называется связь, при которой каждому значению одного признака соответствует:**

*Варианты ответа:*

- а) несколько значений другого, взаимосвязанного с ним признака;
- б) строго определенное значение другого, взаимосвязанного с ним признака.

**23. Корреляционная связь определяется, как связь:**

*Варианты ответа:*

- а) при которой значению каждой величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака;
- б) при которой любому значению одного из признаков соответствует строго определенное значение другого взаимосвязанного с ним признака;
- б) дающая полную характеристику совокупности по ее гомогенности, особенности распределения двух сравниваемых признаков.

**24. Укажите формулу для расчета коэффициента корреляции по Пирсону.**

*Варианты ответа:*

- а)  $1 - \frac{6\sum d^2}{n(n^2 - 1)}$ ;
- б)  $\frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}$ ;
- в)  $\frac{r_{xy}}{r_x}$ .

**25. Корреляционный анализ используется для:**

*Варианты ответа:*

- а) расчета обобщающих коэффициентов, характеризующих различные стороны каждого из изучаемых признаков;
- б) сравнения степени однородности исследуемых совокупностей;
- в) определения пределов возможных колебаний выборочных показателей при данном числе наблюдений;
- г) выявления взаимодействия факторов, определения силы и направленности.

**26. Наиболее точным методом определения степени связи между признаками является метод:**

*Варианты ответа:*

- а) Пирсона;
- б) ранговой корреляции Спирмена.

**27. Достоверным коэффициент корреляции считают только тогда, когда его величина превышает свою среднюю ошибку в:**

*Варианты ответа:*

- а) 3 и более раз;
- б) 2 раза.

**28. Коэффициент корреляции, равный единице, свидетельствует о:**

*Варианты ответа:*

- а) функциональной связи между явлениями;
- б) корреляционной связи между явлениями.

**29. Коэффициент корреляции, равный нулю, свидетельствует о:**

*Варианты ответа:*

- а) отсутствии связи между явлениями;
- б) функциональной связи между явлениями.

**30. Корреляционная связь может быть прямой и:**

*Варианты ответа:*

- а) обратной;
- б) сопряженной.

**31. Величина коэффициента корреляции  $r = 0,6$  свидетельствует о силе между явлениями:**

*Варианты ответа:*

- а) средней;
- б) слабой.

### ОТВЕТЫ К ТЕСТОВОЙ ПРОГРАММЕ

1	<b>б</b>	8	<b>а</b>	15	<b>а</b>	22	<b>б</b>	29	<b>а</b>
2	<b>в</b>	9	<b>б</b>	16	<b>а</b>	23	<b>а</b>	30	<b>а</b>
3	<b>в</b>	10	<b>б</b>	17	<b>а</b>	24	<b>б</b>	31	<b>а</b>
4	<b>б</b>	11	<b>г</b>	18	<b>б</b>	25	<b>г</b>		
5	<b>а</b>	12	<b>а, б, в</b>	19	<b>а</b>	26	<b>а</b>		
6	<b>в</b>	13	<b>а, б, в</b>	20	<b>а</b>	27	<b>а</b>		
7	<b>б</b>	14	<b>а</b>	21	<b>в</b>	28	<b>а</b>		

## ЛИТЕРАТУРА

### ОСНОВНАЯ ЛИТЕРАТУРА

1. *Вальчук, Э. А.* Основы организационно-методической службы и статистического анализа в здравоохранении / Э. А. Вальчук, Н. И Гулицкая, Ф. П. Царук. — Мн., 2003. — 381 с.
2. *Глушанко, В. С.* Общественное здоровье и здравоохранение: курс лекций / В. С. Глушанко. — Витебск, 2001. — 359 с.
3. *Лисицын, Ю. П.* Общественное здоровья и здравоохранении: учебник для вузов / Ю. П. Лисицын. — М.: Издат. дом ГЭОТАР-МЕД, 2002. — 520 с.
4. Общественное здоровье и здравоохранение / Под ред. В. А. Миняева, Н. И. Вишнякова. — М., 2003. — 520 с.

### ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. *Алехнович, М. В.* Методика и организация статистического исследования: учеб.-метод. пос. / М. В. Алехнович. — Гродно, 2001. — 24 с.
2. *Мерков, А. М.* Санитарная статистика / А. М. Мерков, Л. Е. Поляков — Л.: Медицина, 1977. — 384 с.
3. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: учеб. пос. / Под ред. В. З. Кучаренко. — М.: Издат. дом ГЭОТАР-МЕД, 2005. — 188 с.
4. Обучение медицинской статистике: двадцать конспектов лекций и семинаров / Под ред. С. К. Лванга и Чжо-Ек Тыэ. — ВОЗ, Женева: Медицина, 1989. — 215 с.

## СОДЕРЖАНИЕ

Введение .....	3
Цель занятия.....	3
Задачи занятия .....	3
Требования к исходному уровню знаний .....	3
Контрольные вопросы из смежных дисциплин.....	4
Контрольные вопросы по теме занятия .....	4
Учебный материал .....	4
Задания для самостоятельной работы студентов по реализации программных целей занятия .....	15
Литература.....	22

Учебное издание

**Соболева Людмила Григорьевна  
Дорофеев Виталий Михайлович  
Шаршакова Тамара Михайловна**

**ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ОЦЕНКИ  
ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.  
МЕТОД СТАНДАРТИЗАЦИИ. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ**

**Учебно-методическое пособие для студентов 4 курса  
всех факультетов высших медицинских учебных заведений**

**Редактор *Т. Ф. Рулинская*  
Компьютерная верстка *А. М. Елисеева***

Подписано в печать 27.02.2009  
Формат 60 × 84<sup>1</sup>/<sub>16</sub>. Бумага офсетная 65 г/м<sup>2</sup>. Гарнитура «Таймс»  
Усл. печ. л. 1,4. Уч.-изд. 1,5 л. Тираж 150 экз. Заказ № 33

Издатель и полиграфическое исполнение  
Учреждение образования  
«Гомельский государственный медицинский университет»  
246000, г. Гомель, ул. Ланге, 5  
ЛИ № 02330/0133072 от 30.04.2004

