01-Упражнение

Основна терминология. Стандартизация

ас.д-р Костадин Костадинов

Защо учим статистика?

- Статистиката е метод който използваме за да превърнем данните в информация.
- Информацията, която получаваме се превръща в знание.
- Знанието се въвежда в ежедневната медицинска практика, чрез клинични наръчници или политики в общественото здравеопазване.
- По пътя на тази логическа верига, дори и малко нескромно, статистиката е метод (езика на науката), който пряко помага, както на пациента, така и на здравето на обществото.

С какво ще ни помогне статистиката?

- Да взимаме информирани решения в ежедневната ни практика.
- Да взимаме най-добрите решения базирани на доказателства за политики в общественото здравеопазване.
- Да разберем как работи науката.
- За да четем критично нова научна информация.
- За да сме по добри лекари.

Какво няма да научим?

Понастоящем статистиката е силен инструмент в т.н “наука за данните” (data science). В съчетание със сложна математика, програмиране и доза креативност, тази нова дисциплина решава редица практически задачи чрез използване на[1] . В този курс, обаче целта е да придобиете най-основите знания за това как работи статистиката, каква е логиката в нея и какъв език използва.

[1] В момента дори има разработени скенери които “сами разчитат” дали има заболяване и с изчислена възможна “грешка” класифицират какво е то. Това е възможно именно заради инструментите, които статистиката ни предоставя. Разработени са и електрокардиографии записващи сърдечната дейност на пациента и “автоматично” разпознаващи дали е налице определено заболяване.

Терминология

В статистиката един от най-сложните елементи са термините. За да “не сме изгубени в превода” в упражненията, ще въвеждаме тези термини с някои основни примери.

Абсолютни величини

> ⚠ Определение
>
> Това са абсолютни **числа**, които количествено характеризират обемите на статистическите съвкупности или на части от тях, както и значенията на статистическите признаци.

Абсолютните величини са **числа**, които количествено характеризират обемите на статистическите съвкупности или на части от тях, както и значенията на статистическите признаци. Те са винаги **наименовани числа**, измерени в съответните **мерни единици** . Статистическите изследвания обикновено започват с анализ на абсолютни величини, но този тип величини **не са достатъчни** за директни сравнения в **пространствено-времеви аспект.**

Примери за абсолютни величини

Систолното артериално налягане измерено в mmHg е абсолютна величина - има абсолютна стойност, мерна единица

и количествено характеризира систолното артериалното налягане. Кръвната захар измерена в mmol/l също е абсолютна величина - отново е число отразяващо количествено определен признак.

Относителни величини

> ⚠ Определение
>
> Те се изчисляват като частно от делението на две абсолютни величини. Представят се като коефициенти, в проценти (когато коефициентът се умножи по 100), в промили (когато коефициентът се умножи по 1000) и т.н.

Примери за относителни величини

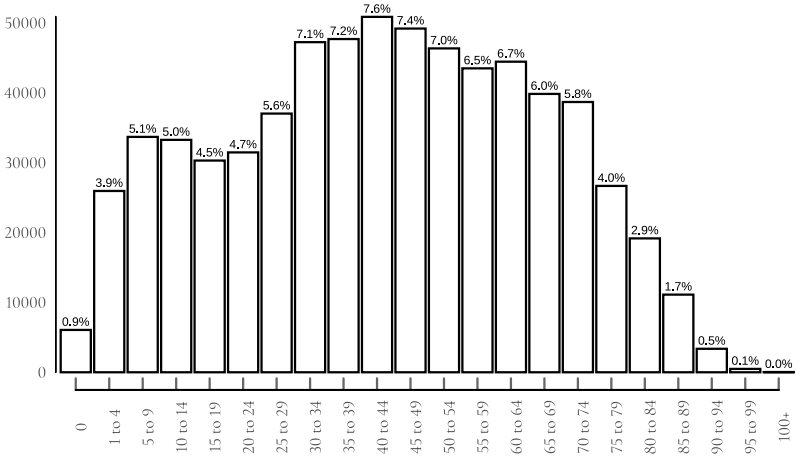
В медицината често използваме относителни величини - когато измерваме помпената функция на сърцето можем да използваме за показател колко милилитра кръв постъпват в аортата след едно сърдечно съкращение - това се нарича ударен обем. Удърният обем е абсолютна величина - има мерна единица (ml) и характеризира сърдечната дейност. Логично е хората с по-висок ръст и тегло (по-едро телосложение) да имат по-високи стойности за този показател отколкото хората с по-нисък ръст и по-малко тегло. Също е логично сърцето на състезател по сумо да изпомпва по-голямо количество кръв (в милилитри) спрямо сърцето на първокласник. Означава ли това, че сърцето на състезателя по сумо работи по-добре от това на първокласника? Отговорът е, че не можем да преценим - двете абсолютни величини не са достатъчни за сравнение. Затова по-важното в случая е какво е съотношението на ударният обем, спрямо количеството кръв налично в сърцето точно преди неговото съкращение. Това е т.н “фракция на изтласкване” и представлява относителна величина. Понеже е отношение на две числа - количеството изтласкана кръв и количеството кръв в сърцето точно преди съкращението.

Екстензивни показатели

> ⚠ Определение
>
> Наричат се още **структурни** и показват как явлението се разпределя на съставните му части, когато то се разглежда самостоятелно само за себе си в определено време и място.

Примери за екстензивни показатели

Ако приемем “възрастта” в гр. Пловдив за статистическо “явление” можем да представим всички жители на град Пловдив в категории според възрастовата им група: новородени до 1г., между 1 и 5 год., от 5 до 10г. и т.н. Ако изчислим броя на хората в съответната възрастовата група спрямо всички жители на града ще получим екстензивен показател. На Фигура 1 е представено разпределението на възрастта в гр. Пловдив. Важно за екстензивните показатели е че сумата от всички елементи е равна на 100%.



Фигура 1: Възраст на населението в гр. Пловдив, 2022г. Пример-екстензивен показател. Графично - ако нанесем броя на хората в определите възрастови групи получаваме т.н разпределение.

Интензивни показатели

> ⚠ Определение
>
> Те се наричат още **честотни** и показват колко често се среща дадено явление в свойствената му среда. Всяка относителна величина е отношение между обемите на две различни статистически съвкупности, но намиращи се във връзка помежду си. В числителя е явлението, а в знаменателя е абсолютният обем на средата, в която възниква определеното събитие.

За да разберем какво е интензивен показател ще дадем един негов представител, често използван в медицината (още по-често неизползван, когато трябва да бъде използван). Това е показателя *леталитет*.

Примери за интензивни показатели

1. **Леталитетът** е показател представяш броя смъртни случаи от конкретно заболяване върху броя на болните от това заболяване в за конкретен период от време и място. Леталитет при заболяването морбили например (дребна шарка) при деца (до 18г.) е 5%. Това означава, че теоретично, на 100 деца със заболяването (дребна шарка) 5 са с летален изход. Тези 5% всъщност показват **честотата** на смъртни случаи при деца болни с дребна шарка- тоест честотата на явлението в неговата свойствена среда. В числителя на този показател постояваме броя смъртни случаи - това е явлението, докато в знаменателя поставяме броя болни деца с морбили- това е обема на средата в която се проявява явлението. На практика числителят и знаменателя представляват две различни статистически съвкупности, въпреки това те има връзка помежду си.
2. **Смъртността** е показател представящ броя на починалите спрямо средния брой население в конкретната област и за конкретно време. Смъртността от конкретно заболяване представлява броя на починалите от заболяването разделен на броя на всички починали в

конкретен период от време и на конкретно място. Двата показателя(обща смъртност и смъртност по причина) не бива да се бъркат с леталитета.

3. **Заболеваемоста**, представлява съотношението на броя новозаболели от някакво заболяване (например от рак на гърдата) спрямо популацията в риск (всички, които биха могли да се разболеят от това заболяване) за даден период от време.

4. В ежедневната практика като лекари също ще ползвате подобни интензивни показатели. Например при пациенти с белодробна астма, вида на използваното лечение зависи от честотата на екзацербации (обостряния) т.н “exacerbation rate”. Това отново е интензивен показател.

Пряк метод на стандартизация

Преди определението за стандартизация, нека започнем с един пример. Нека си представим, че от днес ние сме състава на Министерство на здравеопазването. Вие - министър, а аз ваш съветник със скромна държавна заплата. Изправени сме пред сериозен проблем: В държавата върлува много опасен вирус (много по-опасен от COVID-19). Имаме ваксина, но липсва доверие в нея. Много хора вярват, че ваксините дори убиват. Днес след среща с граждани, противопоставящи се на ваксините, получавате научна статия. Чрез нея, противниците на ваксините се опитват да ви убедят, че ваксините са вредни. В нашият пример приемаме, че този нов вирус може да засегне всички ни еднакво.

В можете да видите данните от тази научна публикация в Таблица 1

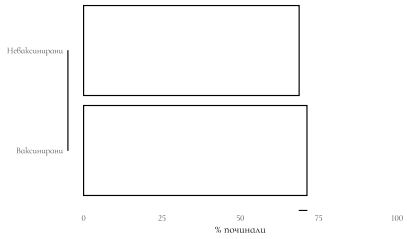
Таблицата е базирана на научно изследване във Великобритания. Авторът посочва, че след 1 година от 4000 ваксинирани са починали 2850 души, докато при неваксинираните от 8000 души са починали 5500. Тези числа всъщност представляват абсолютни величини. За да се опитаме да ги сравним, трябва да използваме относителни величини. Тоест каква пропорция

Таблица 1: Таблица на антиваксарите

	Общо	Починали
Ваксинирани	4000	2850
Неваксинирани	8000	5500

от ваксинираните са починали, спрямо пропорцията на починалите сред неваксинираните. “Сметката” тук е лесна: трябва да разделим броя на починалите ваксинирани върху общият брой ваксинирани, както и броя на починалите неваксинирани, спрямо общият брой неваксинирани. В резултат ще получим относителна величина която също така ще представлява и интензивен показател.

Тук резултатът ни изненадва. Оказва се, че в групата на ваксинираните 71.2% са починали, докато при неваксинираните починали са 68.8%. Това е разлика от 2,4 процентни пункта[2]. Може би наистина “антиваксарите” имат право. Статията изглежда достоверна. Имаме толкова много наблюдавани хора и изглежда, че сред ваксинираните имаме повече починали.



Фигура 2: Разлика между смъртността при ваксинирани и неваксинирани.

[2] Важно: простите аритметични операции между проценти се изразяват в процентни пунктове.

Как бихме могли да си обясним този резултат? Нима наистина ваксините са причина за по-големия брой смъртни случай? Трябва ли да продължим да използваме тази ваксина, ако решението зависеше от нас? Бихме ли посъветвали пациентите си да се ваксинират?

Преди да дадем категоричното си решение, можем да помислим върху данните. Те все още не са информация на която да базираме решенията си. В случая можем да разглеждаме цифрите в таблицата, като сурови данни измерващи една връзка. Тази връзка е между ваксинацията и леталният изход. Не изглежда логично смъртта да се причинява единствено от ваксината или липсата на такава. Има редица други фактори, които влият на смъртта - придружаващи заболявания, предоставената медицинска помощ и може би най-важният сред тях- **възрастта**. Нормално е, ако хората включени в изследването и ваксинирани са по-възрастни то да наблюдаваме и повече починали сред тях.

При сравняването на интензивни статистически показатели се наблюдава фактът, че величината на тези показатели стои в зависимост от структурата на средата, в която изучаваните явления се проявяват. За да проверим дали тази среда “замъглява” връзката между фактора и резултата, можем да използваме статистическия метод на **стандартизацията**.

Например раждаемостта е по-висока в онези населени места, в които преобладава младо население на възраст 20-30 г.

При този и други случаи, когато трябва да се сравняват интензивни статистически показатели, изчислени от среда с различна структура, е необходимо да бъде приложен т.нар. метод на стандартизация.

> ⚠ Определение
>
> Под стандартизация се разбира способът за преобразуване на общите коефициенти, позволяващ да се отстрани или елиминира влиянието на възрастовите или други различия в състава на сравняваните групи.

Стандартизираните показатели позволяват да се анализира и оцени нивото на изучаваното явление при създадени условия на еднородност в сравняваните групи и показват какви биха били общите коефициенти в сравняваните групи, ако тези групи имаха еднакъв състав.

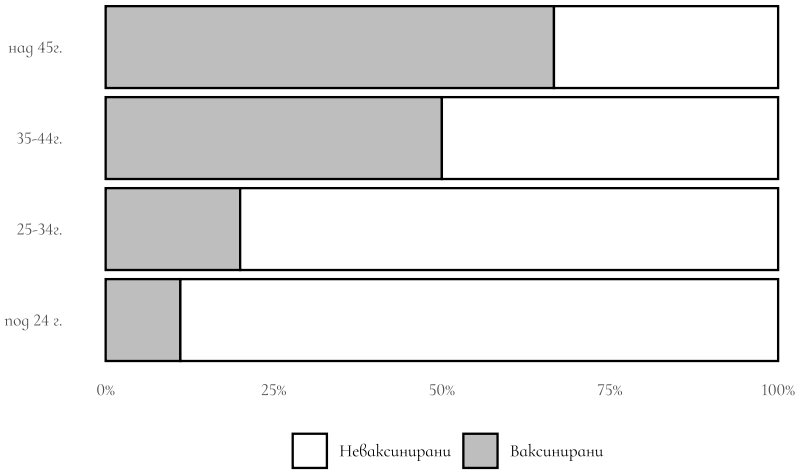
Стъпки

За да извършим стандартизация (в курса по статистика се спираме единствено и само на **прекия метод за стандартизация**) следва да разполагаме с повече данни. Таблицата която разгледахме не съдържа информация за възрастта на участниците. Затова след запитване към авторът на публикацията получаваме по-подробни данни - които можете да видите в Таблица 2.

	Възраст	Общо	Починали
Ваксинирани	под 24 г.	500	250
	25-34г.	500	300
	35-44г.	1000	700
	над 45г.	2000	1600
Неваксинирани	под 24 г.	4000	2400
	25-34г.	2000	1400
	35-44г.	1000	800
	над 45г.	1000	900

Таблица 2: Тазблица с разпределние по възраст

Сега вече имаме повече данни, не само за това колко са били ваксинирани и колко не са, но и в каква възрастова група попадат. Може би, ви прави впечатление от Фигура 3, че ваксинираните са предимно по-възрастни хора, докато при неваксинираните преобладават по-младите.



Фигура 3: Разлика на възрастовата структура между ваксинирани и неваксинирани - възрастовите групи са представени от млади към стари.в групата до 24г. преобладат неваксинирани, докато при над 45-годинишните ваксинираните.

Стъпка 1 Изчисляване на нестандартизираните интензивни показатели

Както по-рано, така и сега, можем да изчислим какъв процент от участниците в двете групи са починали. Това става като разделим броя на починалите върху броя на участниците. В случая ще изчислим този показател за всяка една възрастова група. В Таблица 3 колона леталитет представлява нестандартизираният показател във всяка една от възрастовите групи за ваксинираните и неваксинираните.

	Възраст	Общо	Починали	Леталитет[1]
Ваксинирани	под 24 г.	500	250	0.5
	25-34г.	500	300	0.6
	35-44г.	1000	700	0.7
	над 45г.	2000	1600	0.8
Неваксинирани	под 24 г.	4000	2400	0.6
	25-34г.	2000	1400	0.7

Таблица 3: Нестандартизиран леталитет. Тук е важно да запомним, че общият нестандартизиран показател не е сума от резултатите по подгрупи. Не можем да сумираме тези показатели за ваксинирани и неваксинирани и да ги сравним.

	35-44г.	1000	800	0.8
	над 45г.	1000	900	0.9

[1]Нестандартизиран

Стъпка 2 Изчисляване на "стандарт"

За да можем да сравним двете групи, трябва да "стандартизираме" получените по възрастови групи показатели за леталитета. Само след стандартизация е възможно да сумиране получените числа по възрастови групи и да сравним ваксинирани срещу неваксинирани. За да направим това е необходимо да изберем възрастовата структура една от тези две групи за **стандарт**.

Тук, често възниква въпроса коя структура да изберем за стандарт? Защо да предпочетем едната спрямо другата? Какво е правилото?

Всъщност отговорът на всички тези въпроси е че няма особено значение. Разбира се, че числата след стандартизация ще са различни в зависимост коя структура сме избрали за стандарт, но тук стойността на тези числа не е от толкова голямо значение. От значение е коя от сумарните стойности е по-висока. Независимо коя група сме избрали за стандарт, тази разлика остава една и съща. Независимо коя група изберем за "стандартна" то изводът няма да се промени.[3]

Следващият въпрос, който вероятно възниква е "какво означава да вземем за стандарт структурата на неваксинираните?".

Отговорът е логичен: ако разгледаме само включените участници, които не са ваксинирани можем да изчислим тяхната възрастова структура - ще използваме броя на участниците в определена възрастова група за числител, а общия брой неваксинирани за знаменател. Така след като изчислим какъв процент участници имаме във всяка възрастова група ще получим екстензивен коефициент който ще ползваме за да стандартизираме интензивните показатели за всяка една от възрастовите групи.

В Таблица 4 е изчислен "стандарта" спрямо групата на неваксинираните.

[3] В това упражнение ще докажем това, като извършим стандартизацията, като вземем за стандарт първо структурата на неваксинираните, а после решим същият пример, като вземем за стандарт структурата на ваксинираните.

- Неваксинираните участниците под 24год. са 4000, спрямо общият брой неваксинирани 8000. Стандарта за тази група е 0.5 (ако умножим по 100 ще получим %, т.е. участниците до 24 години) са половината от всички участници. Това ще използваме за стандарт както за ваксинираните така и за неваксинираните.
- По подобен начин е получен и стандарта за възрастовата група от 25 до 34г. В групата на неваксинираните те са 2000, което представлява 25% (или 0.25) от всички неваксинирани. Тази стойност 0.25 ще използваме за стандарт за всички участници от 25 до 34г. - ваксинирани и неваксинирани.
- Това е логиката при всеки един от тези коефициенти в Таблица 4. Може да ви направи впечатление, че сбора на всички стандарти е равен на 1-ца (тоест 100%). Това е така защото този стандарт е екстензивен показател - показва как се разлага явлението "възраст" на съставните и части- отделните възрастови групи.

Възраст	Общо	Починали	Стандарт[1]
под 24 г.	4000	2400	0.500
25-34г.	2000	1400	0.250
35-44г.	1000	800	0.125
над 45г.	1000	900	0.125

Таблица 4: Определяне на стандарт за всяка възрастова група

[1]За изчисляване на колоната стандарт е изпозлвана възрастовата структура на неваксинираните

Стъпка 3: Изчисляване на стандартизираните показатели за леталитет.

След като имаме "стандарт", можем да пристъпим към следващата стъпка. Леталитетът който изчислихме в Таблица 3, беше нестандартизиран. Сега е момента да използваме "стандарта" от стъпка 2 и да го стандартизираме.

Как става това? Решението е лесно: за всяка една от възрастовите групи **умножаваме** нестандартизирания показател по стандарта

Таблица 5: Например, за възрасттовата група до 24 г.при ваксинираните, нестандартизирания леталитет е 0,5, а стандарта 0,250. Стандартизираният леталитет е 0,5x0.250 = 0,125. При неваксинираните, отново във възрастовата група до 24г. нестандартизирания леталитет е 0,6, за да получим стандартизирания умножаваме по стандарта 0,25 - 0,6 x 0.250 = 0.15

	Възраст	Общо	Починали	НС Леталитет[1]	Стандарт[2]	С Леталитет[3]
Ваксинирани	под 24 г.	500	250	0.5	0.500	0.2500
	25-34г.	500	300	0.6	0.250	0.1500
	35-44г.	1000	700	0.7	0.125	0.0875
	над 45г.	2000	1600	0.8	0.125	0.1000
Неваксинирани	под 24 г.	4000	2400	0.6	0.500	0.3000
	25-34г.	2000	1400	0.7	0.250	0.1750
	35-44г.	1000	800	0.8	0.125	0.1000
	над 45г.	1000	900	0.9	0.125	0.1125

[1]Нестандартизиран леталитет
[2]Изчислен спрямо неваксинираните
[3]Стандартизиран леталитет

На този етап стандартизацията е почти изпълнена- налични са стандартизираните показатели като сме умножили нестандартизираните интензивни показатели за леталитет по квотата за стандарт. Тази квота всъщност е изчислена спрямо възрастовата структура на групата неваксинирани (тя беше избрана за стандарт). Вече знаете, че нестандартизираните показатели не се събират. За сметка на това стандартизираните се събират. Когато съберем стандартизираните показатели по възрастови групи ще получим стандартизирания интензивен показател за леталитета - както за ваксинирани така и за неваксинирани. Този показател е стандартизиран спрямо възрастта. Тоест, сме избегнали "замъгляващият ефект" на това, че са ваксинирани предимно възрастни хора, а неваксинирани са преидмно млади.

Стъпка 4: Заключение

Нека извършим тази последна калкулация.

За групата на ваксинираните общият леталитет е:

- **25 %** (стандартизирания леталитет за всички до 24г.) + **15 %** (стандартизирания леталитет за възрастовата група от 25-34г.) + **8,7 %** (стандартизирания леталитет за възрастовата група от 35-44г.) + **10 %** (стандартизирания леталитет за възрастовата група над 45г.). Обшо за всички ваксинирани е стандартизирания леталитет е 58.7%

За групата на неваксинираните общият леталитет е:

- **30 %** (стандартизирания леталитет за всички до 24г.) + **17,5 %** (стандартизирания леталитет за възрастовата група от 25-34г.) + **10 %** (стандартизирания леталитет за възрастовата група от 35-44г.) + **11,25 %** (стандартизирания леталитет за възрастовата група над 45г.). Обшо за всички ваксинирани е стандартизирания леталитет е 65.75%

Заключение

Стандартизираните показатели за леталитет в двете групи са съответно **58.75 %** и **68.75 %**. Леталитетът сред неваксинираните, е с **10 %** по-висок.

Стъпки - при алтернативен избор за стандарт

Предполагам, обаче някои от вас се питат, какво би се случило, ако не ползваме за стандарт възрастовата структура на неваскинираните. Защо избрахме точно нея? Това не е ли пристрастие?

За да докажем, че за изводът няма значение коя структура използваме, ще решим отново примера, като този път за стандарт, използваме структурата на ваксинираните. Отново от стъпка 2.

Стъпка 2 Изчисляваме стандарта (този път спрямо ваксинираните)

В този случай за да изчислим стандарта започваме, като използваме данните само за ваксинираните. В таблицата за ваксинирани установяваме, че участниците под 24г. са 500 от общо 4000. Това означава, че стандарта е 0.125 (или 12.5%). Това извършваме за всяка една от възрастовите групи. Резултатът е видим в Таблица 6 .

Възраст	Общо	Починали	Стандарт[1]
под 24 г.	500	250	0.125
25-34г.	500	300	0.125
35-44г.	1000	700	0.250
над 45г.	2000	1600	0.500

Таблица 6: Определяне на стандарт за всяка възрастова група

[1]За изчисляване на колоната стандарт е изпозлвана възрастовата структура на ваксинираните

Стъпка 3: Изчисляване на стандартизираните показатели за леталитет

След като имаме "стандарт", този път спрямо групата на ваксинираните можем да пристъпим отново към стъпка 3. Сега е момента да използваме "стандарта" от новата стъпка 2. Решението отново е лесно: за всяка една от възрастовите групи умножаваме нестандартизирания показател по стандарта.

Логично, след като сме използвали и друг стандарт, изчислените стандартизирани показатели са различни. Отново напомням, не се интересуваме от конкретното число, а от заключението което ще направим. Отново можем да сумираме стандартизираните показатели.

Таблица 7: Отново, за възрасттовата група до 24 г.при ваксинираните, нестандартизирания леталитет е 0,5, а новия стандарт 0,125. Стандартизираният леталитет е 0,5x0,125 = 0,0625. При неваксинираните, отново във възрастовата група до 24г. нестандартизирания леталитет е 0,6, за да получим стандартизирания умножаваме по стандарта -0,125 - 0,6 x 0,125 =0,075*

	Възраст	Общо	Починали	НС Леталитет[1]	Стандарт[2]	С Леталитет[3]
Ваксинирани	под 24 г.	500	250	0.5	0.125	0.0625
	25-34г.	500	300	0.6	0.125	0.0750
	35-44г.	1000	700	0.7	0.250	0.1750
	над 45г.	2000	1600	0.8	0.500	0.4000
Неваксинирани	под 24 г.	4000	2400	0.6	0.125	0.0750
	25-34г.	2000	1400	0.7	0.125	0.0875
	35-44г.	1000	800	0.8	0.250	0.2000
	над 45г.	1000	900	0.9	0.500	0.4500

[1]Нестандартизиран леталитет
[2]Изчислен спрямо неваксинираните
[3]Стандартизиран леталитет

Заключение

Стандартизираните показатели за леталитет в двете групи са съответно **71.25 %** и **81.25 %**. Леталитетът сред неваксинираните, е с **10 %** по-висок.