

Основни понятия в медицинската статистика. Пряка стандартизация

Статистическа съвкупност, абсолютни и относителни величини, пряк метод на стандартизация — с решена задача стъпка по стъпка

Упражнение 1 · Медицинска статистика · 2026/2027

гл. ас. д-р Костадин Костадинов, д.м.

Катедра по социална медицина и общественото здраве · Медицински университет – Пловдив

Какво ще можете след това упражнение

1. Да обяснявате **защо** лекарят се нуждае от статистика.
2. Да дефинирате **статистическа съвкупност**, **единица на наблюдение** и **признак**; да различавате **генерална** от **извадкова** съвкупност.
3. Да различавате **абсолютни** от **относителни** величини.
4. Да разпознавате и изчислявате **екстензивни**, **интензивни** показатели, показатели за **съотношение** и за **нагледност**.
5. Да обяснявате защо общите интензивни показатели **могат да подвеждат** при сравнение на групи с различен състав.
6. Да извършвате **пряка стандартизация** в четири стъпки и да формулирате извода.

План на часа (90 мин)

1. Увод: защо статистика — 8 мин
2. Статистическа съвкупност и признаци — 12 мин
3. Абсолютни и относителни величини — 26 мин
4. Пряка стандартизация: решена задача — 28 мин
5. Задача в час и домашна работа — 16 мин

Нужни са: калкулатор и листът с формули. Междинните стойности не се закръгляват; крайния резултат закръглявайте до третия знак.

СЪДЪРЖАНИЕ

- 1. Защо учим статистика**
2. Статистическа съвкупност и признаци
3. Абсолютни и относителни величини
4. Пряк метод на стандартизация
5. Задача в час и домашна работа

Защо статистика в медицината

От данни към информация

Статистиката е методът, чрез който **данните** се превръщат в **информация**. „40 починали“ не означава нищо, докато не знаем от колко болни и с какъв състав.

Решения в практиката

Клиничните ръководства, праговете за лечение и здравните политики стъпват на статистически изводи. Лекарят ги прилага всеки ден — често без да го осъзнава.

Език на науката

За да четете **критично** нова научна статия, трябва да разбирате как са получени числата в нея и какво **не** следва от тях.

Целта на курса не е да станете статистици, а да можете да прочетете, проверите и приложите едно число.

№	Тема	Въпрос, на който отговаря
1	Основни понятия. Пряка стандартизация	Как сравняваме честоти в групи с различен състав?
2	Описателна статистика	Как обобщаваме данните — средна, медиана, стандартно отклонение, таблици и графики?
3	Оценка на параметри. t-тест	Какво можем да кажем за популацията по извадка? Различават ли се две средни?
4–5	Асоциация и корелация	Свързани ли са два признака — χ^2 , коефициент на корелация?
6	Регресия	Как да предскажем стойността на един признак по друг?

Всички технологии с „изкуствен интелект“, които ще срещнете в клиниката — от автоматичното разчитане на ЕКГ до анализа на образи — по същество са статистически модели. В този курс учим логиката и езика, на които те стъпват.

СЪДЪРЖАНИЕ

1. Защо учим статистика
- 2. Статистическа съвкупност и признаци**
3. Абсолютни и относителни величини
4. Пряк метод на стандартизация
5. Задача в час и домашна работа

Статистическа съвкупност и единица на наблюдение

Статистическа съвкупност

Множество от **единици на наблюдение**, които са **качествено еднородни** (имат общ признак) и са **ограничени във времето и мястото**.

Единица на наблюдение

Отделният елемент на съвкупността, който се наблюдава и описва: пациент, раждане, смъртен случай, хоспитализация, болница.

Признак

Характеристика на единицата, която се регистрира: пол, възраст, диагноза, систолно налягане, изход от лечението.

Пример

Съвкупност: всички хоспитализации с остър мозъчен инсулт в една болница през 2025 г.

Единица: една хоспитализация.

Признаци: възраст, пол, вид инсулт, продължителност на престоя, изход (жив/починал).

Не е съвкупност

„Всички болни от инсулт“ — няма граници във времето и мястото; не знаем кого броим.

Генерална и извадкова съвкупност

Понятие	Значение
генерална съвкупност (популация)	всички единици, за които искаме да направим извод
извадкова съвкупност (извадка)	частта от генералната съвкупност, която реално наблюдаваме
обем	броят единици: N за популацията, n за извадката
представителност	извадката повтаря структурата на популацията по важните признаци (пол, възраст, тежест на болестта)

Пример

Искаме да знаем колко често се среща хипертония сред възрастните в Пловдив (**генерална съвкупност**). Изследваме 1 200 случайно подбрани жители (**извадка**, $n = 1\,200$).

Внимание

Ако изследваме само пациентите, посетили кардиолог, извадката няма да е представителна: при тях хипертонията е по-честа, отколкото сред всички жители.

Видове признаци

Качествени (атрибутивни)

Описват **свойство**, не количество.

- **номинални** — без естествен ред: пол, кръвна група, диагноза
- **порядкови** — с ред, но без равни интервали: стадий на рака (I–IV), степен на болката (лека, умерена, силна)

Количествени

Изразяват се с **число и мерна единица**.

- **дискретни** — цели числа, получени чрез броене: брой раждания, брой зъби с кариес
- **непрекъснати** — приемат всяка стойност в даден интервал; установяват се чрез измерване: ръст, тегло, кръвна захар, систолно налягане

Количественият признак може да се групира в категории (възраст → възрастови групи). Така работим и днес: стандартизацията използва **възрастови групи**.

СЪДЪРЖАНИЕ

1. Защо учим статистика
2. Статистическа съвкупност и признаци
- 3. Абсолютни и относителни величини**
4. Пряк метод на стандартизация
5. Задача в час и домашна работа

Абсолютни величини

Определение

Числа, които количествено характеризират **обема** на една статистическа съвкупност или на нейни части, както и стойността на конкретен признак.

- Винаги са **наименовани** — имат мерна единица (mmHg, mmol/l, брой случаи, души).
- Получават се чрез пряко **измерване** или **преброяване**.
- Статистическото изследване **започва** с тях.

Примери

- систолно артериално налягане 135 mmHg
- кръвна захар 6,1 mmol/l
- 2 850 починали; 4 000 лекувани
- население на град: 340 000 души

Ограничение

Абсолютните величини **не са достатъчни** за сравнение във времето и пространството. Болница с 400 починали не работи непременно по-зле от болница с 40 — първо трябва да знаем колко болни са преминали през всяка.

Относителни величини

Определение

Получават се при **деление на две абсолютни величини**. Изразяват се като коефициент, процент ($\times 100$), промил ($\times 1\,000$) или на $10\,000 / 100\,000$.

$$\text{относителна величина} = \frac{A}{B} \times k$$

$k = 100$ (%), $1\,000$ (‰), $10\,000$ или $100\,000$

Множителят k се избира така, че числото да е удобно за четене: рядко събитие $\rightarrow$ на $100\,000$.

Клиничен пример: фракция на изтласкване

Ударният обем (ml) е абсолютна величина. Сърцето на едър състезател по сумо изтласква повече милилитри от сърцето на първокласник — това не значи, че работи по-добре.

Фракцията на изтласкване = ударен обем / краен диастолен обем $\times 100$ е относителна величина. Тя прави двете сърца **сравними**.

Четири вида относителни величини

Вид	Какво показва	Знаменател	Пример
екстензивни (структурни)	как едно цяло се разпределя на части	цялото явление	52 % от болните са жени
интензивни (честотни)	колко често възниква явлението в своята среда	средата, в която явлението възниква (под риск)	12 починали на 1 000 души население
съотношение	отношение между две различни по съдържание съвкупности	друга съвкупност (напр. населението)	50 лекари на 10 000 души население
нагледност	относителна промяна на редица спрямо базисна стойност	базисната стойност (= 100 %)	хоспитализациите през 2026 г. са 120 % от тези през 2022 г.

Преди да изчислите или прочетете показател, питайте: **какво е в знаменателя?**

Екстензивни (структурни) показатели

Определение

Показват как едно статистическо явление се **разпределя на съставните си части** в определено време и място.

$$E = \frac{\text{част от явлението}}{\text{цялото явление}} \times 100$$

Свойство

Сборът от всички екстензивни показатели за едно явление е винаги **100 %** (или 1).

Пример: болни от хепатит А в болница Х по възраст

Възраст	< 1	1–4	5–9	10–19	Общо
Болни	70	41	102	87	300
Структура, %	23,3	13,7	34,0	29,0	100,0

$70 / 300 \times 100 = 23,3 \%$ и т.н. Сборът е 100 % (след закръгляне проверявайте с точните стойности).

Интензивни (честотни) показатели

Определение

Показват колко **често** се среща едно явление в **свойствената му среда**. Числителят е броят случаи; знаменателят — средата (лица под риск), в която те възникват.

$$I = \frac{\text{брой случаи на явлението}}{\text{обем на средата}} \times k$$

Показател	Числител	Знаменател
смъртност	починали за година	средногодишно население (‰)
раждаемост	живородени за година	средногодишно население (‰)
заболеваемост	нови случаи за периода	население под риск (на 100 000)
болестност	всички случаи (нови и стари)	население (на 1 000 или %)
леталитет	починали от болестта	болните от болестта (%)
детска смъртност	починали до 1 г.	живородени (‰)

Смъртност ≠ леталитет

При **смъртността** в знаменателя е цялото население; при **леталитета** — само болните. Леталитет 5 % при морбили: от 100 болни деца 5 са починали.

Капан: структурата не е честота

Твърдение в социалните мрежи

„**70 %** от хоспитализираните с инфекцията са ваксинирани. Значи ваксината не помага!“

Това е **екстензивен** показател: описва **състава** на хоспитализираните. Той зависи от това колко души изобщо са ваксинирани — и не казва нищо за **риска**.

Условен пример: град с 100 000 души, 90 % ваксинирани

	Души	Хоспит.	На 100 000
ваксинирани	90 000	70	77,8
неваксинирани	10 000	30	300,0
общо	100 000	100	100,0

Структура на хоспитализираните: 70 % / 30 % (екстензивен). Честота: $300,0 / 77,8 \approx 3,9$ пъти по-висока сред неваксинираните (интензивен).

За сравнение на **риска** са нужни **честотите** в съответните групи; дяловете сред хоспитализираните не са достатъчни.

Показатели за съотношение

Определение

Отношение между две **различни по съдържание** съвкупности. Числителят не е случай на явление сред знаменателя; отношението показва **осигуреност** или **натовареност**.

Използват се най-вече за ресурсите на здравеопазването: лекари, медицински сестри, болнични легла, аптеки — на 1 000 или на 10 000 души население.

Условен пример: област с 680 000 души

- 3 400 лекари: $3\,400 / 680\,000 \times 10\,000 = 50$
лекари на 10 000 души
- 2 720 болнични легла: $2\,720 / 680\,000 \times 1\,000 = 4$
легла на 1 000 души

Не е честота

Не казваме „честотата на лекарите сред населението“. Това е показател за **осигуреност**.

Показатели за нагледност

Определение

Показват **равнището** на членовете на една редица спрямо избрана база (**100 %**). Процентното изменение е показателят минус 100.

- **базисни** — всяка година спрямо една и съща база
- **верижни** — всяка година спрямо предходната

Условен пример: хоспитализации в отделение

Година	2022	2023	2024	2025	2026
Брой	1 250	1 300	1 400	1 150	1 500
Базисни, %	100,0	104,0	112,0	92,0	120,0
Верижни, %	—	104,0	107,7	82,1	130,4

$1\,500 / 1\,250 \times 100 = 120\%$: спрямо 2022 г. хоспитализациите са се увеличили с 20 %.

Бърза проверка: какъв е показателят?

1. 35 % от хоспитализираните са приети по спешност.
2. От 48 болни с рак на устната кухина 12 са починали от него в рамките на една година — 25 %.
3. 4,2 медицински сестри на 1 000 души население.
4. 28 нови случая на рак на гърдата на 100 000 жени за година.
5. Броят на ражданията през 2026 г. е 96 % от този през 2020 г.
6. Болестите на органите на кръвообращението са 60 % от всички смъртни случаи.

Как да решите за секунда

- Знаменателят е **цялото**, разделено на части → **екстензивен**.
- Знаменателят е средата **под риск** → **интензивен**.
- Знаменателят е **друга по съдържание** съвкупност → **съотношение**.
- Знаменателят е **стойност от същата редица** (база) → **нагледност**.

СЪДЪРЖАНИЕ

1. Защо учим статистика
2. Статистическа съвкупност и признаци
3. Абсолютни и относителни величини
- 4. Пряк метод на стандартизация**
5. Задача в час и домашна работа

Задачата: вредна ли е ваксината?

Сценарий

Вие сте министър на здравеопазването. В страната върлува опасна вирусна инфекция. Има ваксина, но гражданите не ѝ вярват. Получавате статия, според която **ваксинираните болни умират по-често**. Данните — за болните с тежка форма, проследени една година:

Група	Болни	Почин.	Леталитет
Ваксинирани	4 000	2 850	71,25 %
Неваксинирани	8 000	5 500	68,75 %

$$2\,850 / 4\,000 \times 100 = \mathbf{71,25\ \%}$$

$$5\,500 / 8\,000 \times 100 = \mathbf{68,75\ \%}$$

Разликата е **2,5 процентни пункта** в полза на неваксинираните.

Въпросът

Спирате ли ваксината? Бихте ли посъветвали пациентите си да не се ваксинират?

Разлика между два процента се изразява в **процентни пунктове**, не в проценти.

Какво „замъглява“ сравнението? Възрастта

Възраст	Ваксинирани			Неваксинирани		
	Болни	Почин.	Структ.	Болни	Почин.	Структ.
< 25 г.	500	250	12,5 %	4 000	2 400	50,0 %
25–34 г.	500	300	12,5 %	2 000	1 400	25,0 %
35–44 г.	1 000	700	25,0 %	1 000	800	12,5 %
≥ 45 г.	2 000	1 600	50,0 %	1 000	900	12,5 %
Общо	4 000	2 850	100 %	8 000	5 500	100 %

Наблюдение

Ваксинираните са **по-възрастни**: половината са на 45 и повече години. Сред неваксинираните половината са под 25 г.

Смесващ фактор

Признак, свързан с **фактора** (ваксинацията) и **независимо от него** — с **резултата** (смъртта), който **не е** последица от фактора. Възрастта отговаря на трите условия: тя може да създаде или да скрие разлика между групите.

Стандартизация: определение

Стандартизацията е преобразуване на общите (сумарните) коефициенти, с което се отстранява влиянието на възрастовите или други различия в състава на сравняваните групи. Стандартизираните показатели показват какви биха били общите коефициенти, ако групите имаха **еднакъв състав**.

— Наръчник по медицинска статистика, гл. 1

Кога я използваме

Когато сравняваме **интензивни** показатели на групи (болници, области, години, държави) с **различна структура** — по възраст, пол, вид заболяване, отделение.

Пряк метод

Изисква **специфичните показатели** по подгрупи за всяка група и **стандартна структура**. Към нея прилагаме специфичните показатели на всяка група. В курса учим само прекия метод.

Пряк метод: четири стъпки

1. **Специфични показатели** P_i за всяка подгрупа и отделно **общият нестандартизиран показател** за всяка сравнявана група.
2. **Стандарт** S_i — структурата на едната група (или на външна популация): $S_i = N_i/N$; $\sum S_i = 1$.
3. **Претеглени приноси по подгрупи** (стандартизирани показатели по подгрупи): $P_i \times S_i$.
4. **Сумиране и заключение**: $P_{\text{ст}} = \sum P_i \times S_i$; сравняваме двата общи стандартизирани показателя.

Формулата в едно

$$P_{\text{ст}} = \sum_{i=1}^k P_i \times S_i$$

P_i — специфичният показател на групата в подгрупа **i**

S_i — делът на подгрупа **i** в стандарта

k — броят подгрупи (тук: 4 възрастови групи)

Ако работите в проценти: $P_i(\%) \times S_i(\%)/100$.

Стъпка 1: специфични (нестандартизирани) показатели

Възраст	Ваксинирани			Неваксинирани		
	Болни	Почин.	P_i	Болни	Почин.	P_i
< 25 г.	500	250	0,500	4 000	2 400	0,600
25–34 г.	500	300	0,600	2 000	1 400	0,700
35–44 г.	1 000	700	0,700	1 000	800	0,800
≥ 45 г.	2 000	1 600	0,800	1 000	900	0,900
Общо	4 000	2 850	0,7125	8 000	5 500	0,6875

$250 / 500 = 0,500$; $2\,400 / 4\,000 = 0,600$ и т.н.

Парадокс

Във **всяка** възрастова група леталитетът на ваксинираните е **по-нисък** с 0,1 (10 процентни пункта). Общо обаче е **по-висок**.

Правило

Общият нестандартизиран показател **не е сбор** от специфичните. Изчислява се от общите числа: $2\,850 / 4\,000$.

Стъпка 2: избор и изчисляване на стандарта

Приемаме за стандарт възрастовата структура на **неваксинираните**:

$$S_i = \frac{N_i}{N} = \frac{4\,000}{8\,000} = 0,500 \quad (< 25 \text{ г.})$$

Възраст	Неваксин.	S_i
< 25 г.	4 000	0,500
25–34 г.	2 000	0,250
35–44 г.	1 000	0,125
≥ 45 г.	1 000	0,125
Общо	8 000	1,000

Стандартът е екстензивен показател

Сборът на S_i е винаги **1** (100 %). Ако не е — върнете се и проверете деленето.

Кой стандарт?

В учебните задачи стандартът е **зададен**. В реални сравнения се предпочита външна популация (напр. Европейската стандартна популация) или сборът от сравняваните групи. Стандартът **винаги се посочва**.

Стъпка 3: претеглени приноси по подгрупи

		Ваксинирани		Неваксинирани	
Възраст	S_i	P_i	$P_i \times S_i$	P_i	$P_i \times S_i$
< 25 г.	0,500	0,500	0,250 0	0,600	0,300 0
25–34 г.	0,250	0,600	0,150 0	0,700	0,175 0
35–44 г.	0,125	0,700	0,087 5	0,800	0,100 0
≥ 45 г.	0,125	0,800	0,100 0	0,900	0,112 5
Общо	1,000	0,712 5	0,587 5	0,687 5	0,687 5

Приносите $P_i \times S_i$ се събират

$$0,250\ 0 + 0,150\ 0 + 0,087\ 5 + 0,100\ 0 = \mathbf{0,587\ 5}$$

Полезна проверка

Групата, чиято структура е стандарт, **запазва** общия си показател: неваксинирани 0,687 5 и преди, и след.

Стъпка 4: заключение

Леталитет	Нестанд.	Станд.
Ваксинирани	71,25 %	58,75 %
Неваксинирани	68,75 %	68,75 %
Разлика	+2,5 п.п.	-10,0 п.п.

Ако двете групи имаха еднаква възрастова структура, леталитетът при ваксинираните би бил с **10 процентни пункта по-нисък**. Изводът на статията се **обръща**.

Как се тълкува стандартизираният показател

- Той **не е реална честота**: никой не е наблюдавал 58,75 % леталитет — наблюдаваният е 71,25 %.
- Отговаря на въпроса „какво **би било**, ако съставът беше еднакъв?“.
- Използва се **само за сравнение** между групи, стандартизирани по един и същ стандарт.
- Не доказва **причинна** връзка — групите може да се различават и по други признаци (хронични болести, достъп до лечение).

Проверка: същата задача с другия стандарт

		Ваксинирани		Неваксинирани	
Възраст	S_i	P_i	$P_i \times S_i$	P_i	$P_i \times S_i$
< 25 г.	0,125	0,500	0,062 5	0,600	0,075 0
25–34 г.	0,125	0,600	0,075 0	0,700	0,087 5
35–44 г.	0,250	0,700	0,175 0	0,800	0,200 0
≥ 45 г.	0,500	0,800	0,400 0	0,900	0,450 0
Общо	1,000		0,712 5		0,812 5

Стандарт: структурата на **ваксинираните** ($500 / 4\,000 = 0,125$ и т.н.).

Резултат

71,25 % срещу 81,25 %: стойностите са **различни**, но посоката е **същата** — по-нисък леталитет при ваксинираните.

Не е универсално правило

Тук изводът не зависи от стандарта, защото разликата е еднаква (10 п.п.) във всяка възрастова група. Ако разликите по подгрупи са с **различна посока**, изборът на стандарт може да промени и извода.

Чести грешки

№	Грешка	Как да я избегнете
1	Сравняване на нестандартизирани показатели при групи с различна структура	Първо погледнете структурата на групите
2	Събиране на специфичните показатели, за да се получи общият нестандартизиран	Общият се изчислява наново от общите числа
3	Събиране вместо умножение в стъпка 3	$P_i \times S_i$ — умножение
4	Стандарт, чийто сбор не е 1 (100 %)	Проверявайте сбора, преди да продължите
5	Тълкуване на стандартизирания показател като реална честота	Той е „какво би било, ако...“
6	Изразяване на разликата между два процента в проценти	Разлика 58,75 % – 68,75 % = –10 процентни пункта

СЪДЪРЖАНИЕ

1. Защо учим статистика
2. Статистическа съвкупност и признаци
3. Абсолютни и относителни величини
4. Пряк метод на стандартизация
- 5. Задача в час и домашна работа**

Задача в час: леталитет в две областни болници

Да се изчислят стандартизираните показатели за леталитет в две областни болници. За стандарт да се приеме **структурата на преминалите болни в болница А**. Закръглете само крайния резултат до третия знак.

Диагноза	Болница А		Болница Б	
	Премин.	Почин.	Премин.	Почин.
рак на стомаха	400	120	200	80
инсулт	500	300	100	70
пневмония	100	40	650	250
Общо	1 000	460	950	400

Въпроси

1. Коя болница има по-висок **нестандартизиран** леталитет?
2. Каква е структурата на болните в болница А?
3. Кои са стандартизираните показатели?
4. Променя ли се изводът? Защо?

Решение

		Болница А		Болница Б	
Диагноза	S_i	P_i	$P_i \times S_i$	P_i	$P_i \times S_i$
рак на сто- маха	0,400	0,300	0,120	0,400	0,160
инсулт	0,500	0,600	0,300	0,700	0,350
пневмония	0,100	0,400	0,040	$\approx 0,385$	$\approx 0,038$
Общо	1,000	0,460	0,460	0,421	0,548

Б, пневмония: $250 / 650 \approx 0,3846$ и $0,3846 \times 0,1 \approx 0,038$; общо Б: $400 / 950 \approx 0,421$; $0,160 + 0,350 + 0,0385 \approx 0,548$.

Отговор

- Нестандартизиран: А **46,0 %**, Б **42,1 %** — Б изглежда по-добре.
- Стандартизиран: А **46,0 %**, Б **54,8 %**.

Б лекува предимно пневмонии (по-нисък леталитет). При структурата на А леталитетът в Б е с **8,8 п.п. по-висок.**

Обобщение и домашна работа

Да запомним

1. Статистическата съвкупност е еднородна и ограничена във времето и мястото.
2. Абсолютните величини не стигат за сравнение; трябва ни **знаменател**.
3. **Екстензивен** — структура (сбор 100 %); **интензивен** — честота; **съотношение** — осигуреност; **нагледност** — динамика.
4. Общите интензивни показатели **могат да подвеждат** при сравнение на групи с **различен състав**.
5. Пряка стандартизация: $P_i \rightarrow S_i \rightarrow P_i \times S_i \rightarrow \sum$ → извод.
6. Стандартизираният показател е **за сравнение**, не реална честота.

Домашна работа

Две задачи по пряка стандартизация — на сайта, раздел „Медицинска статистика“:

kostadinoff.github.io

- Задача 1: смъртност в две общини
- Задача 2: леталитет в две болници

Следващото упражнение: **Описателна статистика** — средни величини, мерки на разсейване, таблици и графики.

Благодаря за вниманието!