**Статистическая оценка факторов заболеваемости туберкулезом в России**

*Бадалов Михаил Александрович, студент 3 курса фельдшерского отделения*

*Ивановский медицинский колледж*

**Аннотация.** Несмотря на периодические изменения в системе подготовки специалистов в медицинской сфере, в программе обучения остаются фундаментальные науки, в том числе математика. Неслучайно в разное время и в разных частях света существовали ученые, которые достигали успеха как в медицине, так и в математике. Ими являются Авиценна (X-XI вв., Азия), Леонардо да Винчи (XV – XVI вв., Европа), Михаил Ломоносов (XVIII в., Россия). В настоящей статье будет рассмотрено практическое применение методов математической статистики для оценки значимости различных факторов заболеваемости туберкулезом в России на основе данных с 2014 по 2022 год. В ходе анализа будут сделаны выводы о целесообразности изменения факторов, имеющих наибольшее влияние на показатель заболеваемости, а также дан прогноз об изменении данного показателя в лучшую сторону с учетом различных рекомендаций.

***Ключевые слова:*** *туберкулез, заболеваемость, коэффициент корреляции, множественная линейная регрессия, статистика.*

**Введение**

Для того, чтобы сделать какие-то выводы, человек, по своей природе, ищет различные взаимосвязи, наблюдает за процессом изменений, строит причинно-следственные связи. Иногда этот процесс занимает достаточно длительное время. Так, например, долгое время курение табака было модным в обществе и считалось полезным, до тех пор, пока люди не стали замечать взаимосвязь болезней легких с курением. Это привело к тому, что отношение к данному процессу изменилось, став негативным. Причина длительности данных наблюдений состоит в том, что отсутствовала какая-либо четкая структура исследований, что заставило людей не изменять данной привычке приблизительно 200 лет.

На основе наблюдений и взаимосвязи различных факторов был открыт вирус клещевого энцефалита. Он был открыт в 1937 году советским ученым Л. Зильбером в кратчайшие сроки после того, как учёный понял взаимосвязь увеличения заболеваемости и сезона. Клещевой энцефалит имеет сезонный характер, соответствующий активности клещей. Максимальный подъем заболеваемости наблюдается в мае — июне. Второй, менее выраженный подъём заболеваемости, регистрируется в августе — сентябре. Данное наблюдение позволило направить вектор изучения в нужное русло, что позволило выделить вирус в течение трех месяцев, хотя до этого долгое время вирусологи не могли понять причину и первоисточник этой опасной болезни [1].

В настоящее время, с учетом развития интернета, информация стала намного доступнее, чем была в прошлом, что позволяет сделать необходимые выводы быстрее. Кроме того, с учетом развития математической науки, появились точные и измеряемые алгоритмы анализа, чтобы избежать субъективных оценок и «прикидки на глаз».

**Корреляция**

Статистическая взаимосвязь двух и более величин, в ходе которой, при некотором изменении одной из них другая может тоже измениться, называется корреляционной зависимостью, или корреляцией. Для оценки этой зависимости используется такой показатель как коэффициент корреляции – статистическая мера, которая вычисляет силу связи между относительными движениями двух переменных. Принимает значения [-1, 1] и показывает силу и направление взаимосвязи двух количественных переменных. Знак коэффициента корреляции показывает направление взаимосвязи – если он положительный, то зависимость прямая, то есть с увеличением одной величины увеличивается и другая, если отрицательный, то зависимость обратная (с увеличением одной величины другая уменьшается). Чем больше модуль коэффициента, тем сильнее зависимость. Родоначальниками коэффициента корреляции являются французский физик Огюст Браве и английский ученый Фрэнсис Гальтон. В 1844 году Огюст Браве опубликовал статью о статистической концепции корреляции и пришел к определению коэффициента корреляции [4].

Существует несколько способов расчета коэффициента корреляции. Самым распространенным является расчет по Пирсону, или, как его еще называют, метод квадратов. Именно он и будет использоваться для дальнейших расчетов.

Критерий учета корреляции Пирсона был разработан командой британских ученых во главе с Карлом Пирсоном (1857-1936) в 90-х годах 19-го века, и поэтому данный метод назван его именем. Основным преимуществом является то, что при расчетах не имеет значение, в каких единицах измерения представлены исследуемые величины. Основным недостатком следует назвать недостаточно качественный учет значений, явно выделяющихся из общей тенденции, например, так называемую «ошибку выжившего», когда человек, ведущий нездоровый образ жизни, доживает до 100 и более лет, и попадает в статистический анализ. Поэтому, имея данные значения в анализируемом массиве данных, нужно принимать их во внимание. При этом следует понимать, что чем больше данных имеется для анализа, тем более полная получится итоговая картина [3].

Формула коэффициента корреляции по методу Пирсона выглядит следующим образом [5]:

$$r=\frac{∑\left(x-\overbar{x}\right)⋅\left(y-\overbar{y}\right)}{\sqrt{∑\left(x-\overbar{x}\right)^{2}Σ\left(y-\overbar{y}\right)^{2}}}$$

В практической части, особенно с учетом наличия множества необходимых показателей для анализа, в современности используются компьютерные методы вычисления. Одним из которых является Microsoft Excel.

**Анализ факторов заболеваемости туберкулезом**

Заболеваемость – это количество впервые выявленных случаев болезни на 100 тысяч населения. Для анализа влияния различных факторов следует отобрать наиболее важные предположения для того, чтобы их подтвердить или опровергнуть. Туберкулез является социально значимым заболеванием, поэтому следует учесть такие параметры, как:

* наличие контакта с больными туберкулезом людьми или животными;
* наличие социальной дезадаптации;
* неудовлетворительные условия труда и быта;
* неполноценное питание;
* алкоголизм;
* табакокурение;
* употребление наркотиков;
* ВИЧ-инфицированность
* наличие сопутствующих заболеваний (сахарный диабет, язвенная болезнь желудка и 12-ти перстной кишки, хронические неспецифические болезни легких, иммунодефицитные состояния и заболевания). [2]

Особенностью математических методов является строгая точность, поэтому в ходе анализа все показатели должны быть конкретными и исчисляемыми. При этом, как уже уточнялось, не особо важно смотреть на единицы измерения, поэтому для оценки факторов риска можно использовать данные из открытого доступа. Для определения факторов риска выделяются следующие показатели, которые будут использованы в расчетах:

* Основной показатель – заболеваемость туберкулезом (случаев на 100 тыс. населения)
1. Количество рожденных детей (млн чел.)
2. Розничные продажи сигарет (тыс./ч.)
3. Лица, находящиеся в местах лишения свободы (тыс. чел.)
4. Средняя заработная плата (номинал, р.)
5. Затраты на здравоохранение (трлн. р.)
6. Лица, прошедшие диспансеризацию (млн. ч.)
7. Обеспеченность жильем (м2/ч.)
8. Учтенный сахарный диабет (люди на учете на 100 тыс. населения)
9. Заболеваемость ВИЧ (случаев на 100 тыс. населения)
10. Употребление алкоголя (Л/ч. чистого спирта)

Данные показатели позволяют дать однозначные данные по интересующему нас периоду времени. Например, уровень курения отлично показывают розничные продажи сигарет, поскольку тяжело в ходе анкетирования или опроса определить данный показатель, т.к. кто-то курит больше, кто-то меньше. Основным источником данных является официальный сайт российской статистики [7], некоторые данные были взяты из СМИ [8][9][10][11]. Данные по данным показателям с 2014 по 2022 год представлены в таблице, где также рассчитан коэффициент корреляции.

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Туб. заб.** | **Год** | **Рождение** | **Сигареты** | **МЛС** | **ЗП** | **Затраты трлн р.** | **Дисп. млн.** | **Жилье** | **СД** | **ВИЧ** | **Алкоголь** |
| 62,2 | 2014 | 1,942 | 1,73 | 677,3 | 32495 | 2,5000 | 39,6 | 23,7 | 165,0 | 3,3 | 8,85 |
| 57,5 | 2015 | 1,940 | 1,62 | 639,9 | 34030 | 2,8610 | 63,0 | 24,4 | 166,7 | 3,3 | 8,41 |
| 53,0 | 2016 | 1,889 | 1,70 | 630,1 | 36709 | 2,8130 | 19,0 | 24,9 | 167,8 | 2,8 | 8,42 |
| 48,0 | 2017 | 1,690 | 1,65 | 602,2 | 39167 | 3,0350 | 26,1 | 25,2 | 170,3 | 2,6 | 7,20 |
| 29,6 | 2018 | 1,604 | 1,60 | 563,2 | 43724 | 3,3159 | 21,9 | 25,8 | 174,3 | 2,4 | 7,30 |
| 40,9 | 2019 | 1,481 | 1,50 | 410,0 | 47867 | 3,7897 | 40,8 | 26,3 | 177,4 | 2,1 | 9,06 |
| 32,1 | 2020 | 1,437 | 1,52 | 384,0 | 51344 | 4,9393 | 38,4 | 26,9 | 180,3 | 2,5 | 9,13 |
| 30,9 | 2021 | 1,398 | 1,48 | 357,0 | 57244 | 4,4000 | 50,0 | 27,8 | 185,2 | 2,0 | 10,30 |
| 31,1 | 2022 | 1,306 | 1,46 | 266,0 | 65338 | 1,2430 | 68,0 | 28,2 | 191,0 | 1,7 | 8,62 |
| **Коэф. Кор.** |  | **0,91** | **0,82** | **0,81** | **-0,85** | **-0,33** | **-0,08** | **-0,89** | **-0,85** | **0,86** | **-0,18** |

Согласно общепринятым критериям, статистическая взаимосвязь признается значимой при коэффициенте корреляции более 0,7, средней – при значении от 0,3 до 0,7, незначительной менее 0,3 по модулю. Знак указывает прямую или обратную зависимость. В ходе расчета значимыми прямыми показателями оказались:

* Количество рожденных детей (значение 0,91)
* Розничные продажи сигарет (0,82)
* Лица, находящиеся в местах лишения свободы (0,81)
* Заболеваемость ВИЧ (0,86)

Значимыми обратными показателями оказались:

* Номинальная заработная плата (-0,85)
* Обеспеченность жильем (-0,89)

В целом, это позволяет подтвердить гипотезу о важности основных критерием, что, в итоге, помогает дать рекомендации по коррекции показателя в лучшую сторону с целью снижения заболеваемости туберкулезом. Из общей картины выпадает связь туберкулеза с сахарным диабетом, в ходе анализа оказалось, что имеется явная обратная зависимость, что противоречит основной гипотезе (коэффициент -0,85). В таком случае следует изучить причину такого результата. Скорее всего, связь оказалась обратной из-за неправильной интерпретации или кодировки статистического показателя – заболеваемость туберкулезом считалась по первично зарегистрированным случаям, в то время как по сахарному диабету учитывались все лица, состоящие на учете по данной патологии. В то же время, показатель ВИЧ как сопутствующего заболевания дал правильный результат, поскольку параметр был аналогичным основному исследуемому.

Не подтвердилась взаимосвязь с заболеваемостью туберкулезом таких показателей, как затраты государства на здравоохранение, прохождение диспансеризации населением и употребление алкоголя. Предполагается, что затраты показали незначительную отрицательную корреляцию в связи с тем, что в период исследования включилась пандемия COVID-19, что потребовало крупных затрат в сфере здравоохранения. Прохождение диспансеризации и употребление алкоголя не показали значимой взаимосвязи с заболеваемостью туберкулезом, и это можно связать с субъективностью данных показателей.

**Прогнозирование с помощью множественной линейной регрессии**

Проведя анализ значимости факторов заболеваемости туберкулезом и имея понимание о самых значимых, можно составить уравнение множественной линейной регрессии и дать рекомендации по снижению данного показателя. Зачастую на практике значения переменных определяются влиянием не одного, а нескольких факторов, что и показал анализ при расчете коэффициента корреляции. При наличии многих значимых факторов каждый из них будет иметь свой особенный вес в итоговой модели. Модель множественной регрессии – это модель зависимости результирующей переменной более чем от

одной независимой переменной.В общем виде линейной уравнение множественной регрессии можно записать следующим образом:

у = a1x1 + а2х2 + ... +anxn + b + £

где

у - результативный признак (зависимая, результирующая, эндогенная переменная);

п - число факторов, включенных в модель;

х1, х2, ..., хn - признаки-факторы (регрессоры, объясняющие, предикторные, предопределенные, экзогенные переменные);

а1 ,а2, ..., аn - коэффициенты регрессии;

b - свободный член регрессии;

£ - компонента, отражающая в модели влияние случайных факторов, из-за которых реальное значение показателя может отклоняться от теоретического (регрессионный остаток). [6]

Для составления уравнения множественной линейной регрессии в данном исследовании следует использовать программное обеспечение. При этом следует учитывать только те показатели, которые являются значимыми, исходя из расчета коэффициента корреляции. В ходе анализа с помощью Microsoft Excel получились результаты расчета, представленные в таблице.

Таблица 2

|  |  |  |
| --- | --- | --- |
| **Год** | **Расчетный показатель** | **Реальный показатель** |
| 2014 | 62,2 | 62,2 |
| 2015 | 56,7 | 57,5 |
| 2016 | 55,2 | 53,0 |
| 2017 | 41,6 | 48,0 |
| 2018 | 34,9 | 29,6 |
| 2019 | 41,4 | 40,9 |
| 2020 | 33,8 | 32,1 |
| 2021 | 28,8 | 30,9 |
| 2022 | 30,7 | 31,1 |

Полученные результаты позволяют сделать вывод о том, что уравнение линейной регрессии получилось истинным. Это позволяет сделать прогноз на следующий период по заболеваемости туберкулезом в случае изменения значимых факторов в лучшую сторону.

Таблица 3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Год** | **Дети родились** | **Продажи сиг.** | **МЛС тыс.** | **ЗП** | **Затраты трлн р.** | **Дисп. млн.** | **жилье** | **СД** | **ВИЧ** | **Алкоголь** |
| 2022 | 1,306 | 1,46 | 266,0 | 65338 | 1,2430 | 68,0 | 28,2 | 191,0 | 1,7 | 8,62 |
| прогноз | 1,50 | 1,4 | 250 | 70000 | 2 | 70 | 30 | 191,0 | 1,5 | 8 |
| Улучшение, % | 14,8% | 4,1% | 6,0% | 7,1% | 60,9% | 2,9% | 6,4% | 0,0% | 11,8% | 7,2% |

Согласно составленному уравнению множественной линейной регрессии, при улучшении исследуемых параметров согласно приведенной таблице, показатель заболеваемости туберкулезом в России составит 18,3 (человека на 100 тыс. населения), что на 41,15% меньше, чем последнее реальное зарегистрированное значение.

**Вывод**

Показанные в статье методы могут быть применимы для анализа в сфере здравоохранения, поскольку могут подтвердить или опровергнуть некоторые предположения относительно влияние тех или иных показателей на итоговый результат по различным заболеваниям. Такое социально значимое заболевание, как туберкулез, имеет много различных факторов заболеваемости. Проведя анализ данных с 2014 по 2022 год с помощью таких методов математической статистики как расчет коэффициента корреляции и построение уравнения множественной линейной регрессии, удалось определить основные показатели, сопутствующие показателю заболеваемости туберкулезом. Прямую зависимость показывают: количество рожденных детей, розничные продажи сигарет, лица, находящиеся в местах лишения свободы, заболеваемость ВИЧ. Обратную зависимость показывают: номинальная заработная плата, обеспеченность жильем. Данное исследование позволяет дать рекомендации по дальнейшему снижению показателя заболеваемости. Так, в случае увеличения рождаемости на 14,8%, заработной платы на 7,1%, обеспеченности жильем на 6,4%, а также снижения заболеваемости ВИЧ на 11,8%, числа заключенных на 6%, продажи сигарет на 4,1% следует ожидать снижения заболеваемости туберкулезом на 41,15%.

**Список литературы**

1. 75-летие открытия вируса клещевого энцефалита. Сравнение ранних (1937-1945) и современных штаммов - Погодина В.В., Карань Л.С., Левина Л.С., Колясникова Н.М., Герасимов С.Г., Маленко Г.В. -, ФГБУ «Институт полиомиелита и вирусных энцефалитов им. М.П. Чумакова» РАМН; ЦНИИ эпидемиологии Роспотребнадзора РФ;
2. Профилактика заболеваний - Алексеенко С. Н., Дробот Е. В., глава 19, ТУБЕРКУЛЕЗ: ЭПИДЕМИОЛОГИЯ, ФАКТОРЫ РИСКА, ПРОФИЛАКТИКА, Издательство: Академия Естествознания, 2015;
3. Коэффициент корреляции Пирсона - Мирослав Меньшов, - Кафедра математической статистики ИВМиИТ КФУ, 2020;
4. Общая теория статистики: Учебник / Под ред. Р. А. Шмойловой. — 3-е издание, переработанное. — Москва: Финансы и Статистика, 2002;
5. Елисеева И. И., Юзбашев М. М. Общая теория статистики: Учебник / Под ред. И. И. Елисеевой. — 4-е издание, переработанное и дополненное. — Москва: Финансы и Статистика, 2002;
6. ЭКОНОМЕТРИКА Часть 2: Множественная регрессия и корреляция в эконометрических исследованиях, учебное пособие - Н.М. Удинцова, Н.А. Коптева, ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ДОНСКОЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ», кафедра высшей математики, Зерноград, 2016
7. <https://rosstat.gov.ru/> - официальный сайт Федеральной службы государственной статистики, электронный ресурс
8. <https://aif.ru/> - электронный ресурс
9. <https://wciom.ru/> - электронный ресурс
10. <https://lenta.ru/> - электронный ресурс
11. <https://tass.ru/> - электронный ресурс