

Военно-медицинская академия имени С.М.Кирова
Кафедра (автоматизации управления медицинской службой
с военно-медицинской статистикой)

«УТВЕРЖДАЮ»
Заведующий кафедрой
(автоматизации управления медицинской
службой с военно-медицинской статистикой)
доктор медицинских наук профессор

А. Корнеенков

« ____ » _____ 2022 г.

Преподаватель кандидат медицинских наук доцент А.Кобзев.

ЛЕКЦИЯ № 3

по дисциплине «МЕДИЦИНСКАЯ СТАТИСТИКА И ИНФОРМАТИКА»

на тему: «СТАТИСТИЧЕСКАЯ ОЦЕНКА СВЯЗЕЙ МЕЖДУ
ПРИЗНАКАМИ»

для слушателей клинической ординатуры
клинического и медико-профилактического профиля

Обсуждена и одобрена на заседании кафедры
Протокол № 11
« 22 » ноября 2022 г.

Уточнено (дополнено):

« ____ » _____ 20 ____ г. _____

СОДЕРЖАНИЕ

№ п/п	Учебные вопросы	Время, мин
1	Введение	5
2	1. Виды связей между признаками	20
3	2. Корреляционный анализ	25
4	3. Регрессионный анализ	35
5	Заключение	5

Учебное время – 2 часа.

Литература

а) Использованная для подготовки текста лекции:

1. Военно-медицинская статистика: учебник; под ред. В.И. Кувакина, В.В. Иванова. – СПб.: ВМедА, 2005. -528 с.
2. Лядов В.Р. Основы теории вероятностей и математической статистики. Для студентов медицинских ВУЗов // Информационно-аналитическая библиотека. - Вып.2.-СПб., 1998.-108 с.
3. Математико-статистические методы в клинической практике. Учебное пособие. - СПб.: ВМедА, 1993.

б) Рекомендуемая обучаемым для самостоятельной работы по теме лекции:

4. Военно-медицинская статистика: учебник; под ред. В.И. Кувакина, В.В. Иванова. – СПб.: ВМедА, 2005. -528 с.
5. Математико-статистические методы в клинической практике. Учебное пособие. - СПб.: ВМедА, 1993.

НАГЛЯДНЫЕ ПОСОБИЯ

1. Мультимедийная презентация.

ТЕХНИЧЕСКИЕ СРЕДСТВА ОБУЧЕНИЯ

1. Персональный компьютер (ноутбук).
2. Мультимедийный проектор.

Текст лекции
(СЛАЙД 1, СЛАЙД 2)

Рекомендованная по данной теме литература представлена на **СЛАЙДЕ 3.**

ВВЕДЕНИЕ

В практике работы различных должностных лиц и специалистов военно-медицинской службы, при проведении статистического анализа данных, наряду с ответом на вопрос о достоверности различий сравниваемых числовых характеристик тех или иных признаков или явлений достаточно часто возникает потребность в ответе на другой важный вопрос. А именно: "Оказывает ли влияние величина одного признака на значения (величину) какого-либо другого или нескольких других признаков?". Другими словами - "Существует ли связь, взаимное влияние между какими-либо признаками?".

Следует всегда помнить, что все явления в природе и обществе и весь мир в целом представляют собой сложную систему отношений, существенной стороной которой является **диалектическая связь и взаимодействие причин и следствий**. Благодаря этой связи, одни явления и процессы порождают другие, осуществляется вечное движение и развитие. Общая закономерность связи и взаимодействия относится и к случайным явлениям, изучаемым и военной медициной. Так, имеется определенная связь между признаками физического развития военнослужащих (длина, масса тела и др.), между признаками, характеризующими функционирование систем здорового организма и при патологии (температура тела, частота пульса и др.). Для того чтобы правильно понять любое явление, нужно рассматривать его в связи с другими явлениями. Методы изучения связей таких случайных явлений, как условия воинского труда и быта, - с одной стороны, и показателями состояния здоровья и заболеваемости воинского коллектива, - с другой, занимает важное место в военно-медицинской статистике. Врач должен быть вооружен методами для вскрытия связи между

случайными явлениями и методами их количественной оценки.

Исследованиям связей между признаками, количественным оценкам направления и силы такой связи посвящены специальные разделы прикладной статистики, которые носят название "Корреляционный анализ" и "Регрессионный анализ", которые мы с вами сегодня рассмотрим.

Вопрос 1. Виды связей между признаками

В соответствии с основными положениями теории вероятностей в природе различают два вида связи - связь между случайными величинами может быть *функциональной и статистической*.

Сразу условимся, что в лекции будем рассматривать простой случай, когда связь изучается между двумя признаками, выраженными количественно. Эти признаки в общем виде принято обозначать буквами латинского алфавита - X и Y , причем Y является некоторой функцией от X , т.е. $Y = f(X)$. Поскольку значения признака Y зависят от значений признака X , признак X принято называть признаком-причина, а Y - признаком-следствие.

Функциональная связь. (СЛАЙД 4) *Функциональной* называется такая связь между признаками X и Y , когда каждому допустимому значению признака X по определенному закону соответствует единственное и строго определенное значение признака Y .

Функциональная связь является строгой, точной, полной зависимостью, которая, как правило, может быть выражена в виде уравнения. Чаще всего она наблюдается и изучается в точных науках, главным образом в математике и физике.

Все мы помним раздел математики, который изучали в средней школе – «Анализ функций», где как раз и изучались функциональные связи, т.е. зависимости между двумя переменными (аргумент и функция), описываемые в виде уравнений или функций, например, квадратичная функция – $Y=x^2$, отображаемая графически в виде параболы. $Y=x^3$ – парабола третьего порядка. В

«Геометрии» - формулы расчета площадей и объёмов фигур, например площади круга – $S_{\text{кр}} = \pi R^2$.

В физике напомним, нами изучались различные законы, в частности «газовые законы» (Гей-Люссака, Бойля-Мариотта), описывавшими зависимость между различными переменными в виде уравнений. Например, зависимость объема газа от температуры, парциального давления кислорода от высоты над уровнем моря, температуры кипения воды от высоты над уровнем моря, между толщиной свинцового экрана и доли поглощаемого им гамма-излучения и др.

Особенностью функциональных связей является то, что нам обычно известен полный перечень факторов, определяющих величину результирующего признака, а также механизм этого влияния.

В медицинских и биологических науках, там, где предметом изучения являются объекты живой природы и прежде всего - человек, наблюдаются более сложные взаимоотношения между различными признаками и их влияние друг на друга. Это прежде всего связано с тем, что живые организмы, особенно человеческий, представляют собой гиперсложные саморегулирующиеся иерархические системы. Каждый организм уникален и может по-разному реагировать на то или иное воздействие. Здесь мы чаще всего сталкиваемся с другим видом связи между теми или иными признаками - с корреляционной, или вероятностной связью, относящихся к статистической связи.

(СЛАЙД 5) Статистической, т.е. вероятностной, корреляционной называется такая связь между признаками X и Y , когда строго определенному значению признака X соответствует некоторое множество возможных значений признака Y .

Приведём примеры. Так, например, известна зависимость (т.е. связь) между температурой тела и частотой пульса (ЧСС), особенно у больных с заболеваниями, сопровождающимися лихорадкой. Известно так же, что увеличение сроков доставки раненых с проникающими ранениями живота в

хирургический стационар ведет к увеличению уровня летальности. Другой пример, с понижением температуры наружного воздуха в осенне-зимний период возрастает число простудных заболеваний, а в летний период при повышении температуры – кишечных инфекций. Почему эти связи являются вероятностными, корреляционными? (Проводится разбор примеров. Затем у аудитории просим привести другие медицинские примеры...).

Как в случае функциональных, так и корреляционных связей между признаками можно и принято говорить о **направлении** этих связей. (СЛАЙД 6) Если при функциональной или корреляционной связи изменения величины факторного и результативного признака совпадают, то есть с увеличением значений первого признака X , увеличиваются значения второго признака Y и наоборот, то говорят о наличии **прямой, или положительной связи**.

Если функциональные или корреляционные связи характеризуются несовпадением направлений изменения факторного и результативного признаков, то говорят о наличии **обратной, или отрицательной связи** между ними.

По аналитическому выражению корреляционная связь может быть прямолинейной (линейной) и криволинейной (нелинейной). Прямолинейной называется связь, когда величина результативного признака изменяется приблизительно равномерно с изменением величины влияющего фактора, а криволинейной называется связь, при которой величина результативного признака неравномерно возрастает и убывает с изменением величины влияющего фактора. Математически прямолинейная связь между факторными и результативными признаками может быть выражена в виде уравнения прямой, а при криволинейной — в виде уравнения кривой, например, гиперболы, параболы, показательной функции и т.п.

Следует четко различать понятия прямой и прямолинейной связи, так как они характеризуют различные стороны этого явления. В первом случае речь идет о направлении связи, во втором — о форме ее аналитического выражения. При прямолинейной связи может быть и обратная зависимость.

Поскольку функциональные связи являются полными и строгими, то они точно описываются определенным аналитическим выражением. Корреляционные же связи, будучи неполными, нестрогими, выражаются определенным уравнением лишь приблизительно.

При классификации взаимосвязей учитывается число факторных признаков, оказывающих влияние на результативный показатель. Если исследуется зависимость результативного признака от какого-либо одного фактора, то имеет место простая однофакторная связь. Но о ней в полной мере можно говорить только в случае функциональной связи. По иному следует рассматривать однофакторную корреляционную связь. При корреляционных вероятностных зависимостях на результативный признак в действительности влияет множество факторов. Следовательно, если мы оцениваем только один фактор, то влияние других просто не учитываем, абстрагируемся от них. Подобная корреляция называется парной. Хотя, строго говоря, корреляционный анализ позволяет определить степень влияния на выходной параметр нескольких факторов.

В практике статистических исследований нередко возникает необходимость в изучении и измерении связи между результативным признаком и двумя и более факторными, то есть рассматривается *множественная* корреляция. Она решает три задачи: 1) определяет форму связи, 2) тесноту связи, 3) влияние отдельных факторов на общий результат.

Рассмотрим методы, используемые в статистике для оценки связи между признаками.

Методы статистической оценки связи между признаками

Для выявления и оценки связи между признаками в медицинской статистике существует несколько методов, основными из которых являются (СЛАЙД 7):

- Анализ с помощью диаграмм;
- Корреляционный анализ;
- Регрессионный анализ.

Одним из вариантов диаграмм является **Диаграмма "причина-следствие" (СЛАЙД 8).**

Диаграмму "причина-следствие" иногда называют по фамилии автора диаграммой Исикавы или, по внешнему виду, "рыбий скелет".

Диаграмма "причина-следствие" строится следующим образом:

- сформулировать проблему ("голова рыбы"), которую записать во главе горизонтальной прямой ("хребет"),
- записать наиболее существенные факторы и условия, влияющие на суть проблемы, в начале больших наклонных линий ("большие кости"),
- нанести совокупность причин, влияющих на наиболее существенные факторы и условия, на мелкие линии ("средние и мелкие кости"),
- факторы и условия ранжировать по значимости,
- сформулировать статистические гипотезы о связях случайных величин.

Другим видом диаграмм являются **Диаграммы рассеивания (СЛАЙД 9).**

Алгоритм построения *диаграммы рассеивания*:

- выбрать и назвать случайные величины – составляющие системы,
- найти минимум и максимум каждой из них,
- построить оси координат, нанести на них масштабную сетку,
- нанести на график результаты экспериментов (точки),
- проанализировать наличие, вид и силу статистической связи между случайными величинами.

Для избежания ошибок координатные оси должны быть равновелики,

следует учитывать возможность наличия в замерах выбросов, промахов...

По характеру расположения точек на диаграмме рассеивания можно сделать предварительный вывод о форме (линейная – криволинейная), направлении (положительная – отрицательная) и силе связи (сильно рассеянное облако – концентрированное вокруг некоторой тенденции). Беспорядочное расположение точек на графике указывает на отсутствие связи между признаками (рис.2).

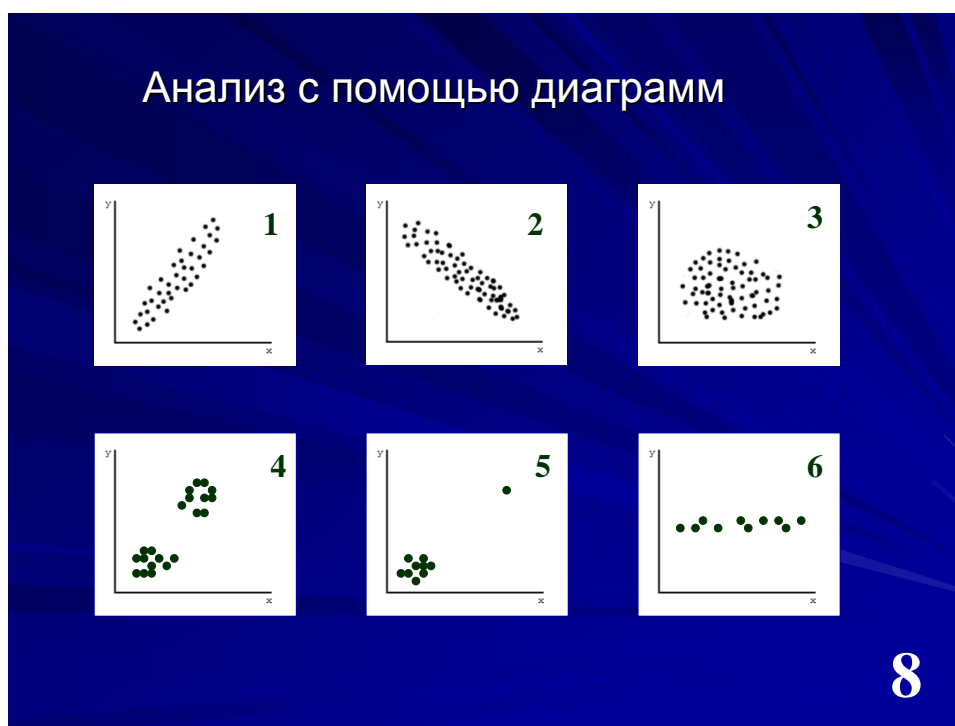


Рис. 2. Анализ с помощью диаграмм

Однако наиболее точно, количественно охарактеризовать связи между признаками позволяют специально разработанные методы статистического анализа.

Вопрос 2. Корреляционный анализ

(СЛАЙД 10) Целью корреляционного анализа является: установление наличия или отсутствия статистически значимой корреляционной связи между признаками, определение направления и силы этой связи.

Если в результате анализа диаграммы рассеивания можно предположить наличие прямолинейной корреляционной связи, то в этом случае следует приступить к вычислению **линейного коэффициента корреляции**.

Линейный коэффициент корреляции вычисляют для измерения силы связи между двумя признаками, если известно, что зависимость одного из признаков от другого близка к прямолинейной. Кроме этого, данные признаки должны быть измерены в интервальной шкале или шкале отношений, т.е. быть **количественными**. Впервые данный коэффициент корреляции был предложен Пирсоном. В соответствии с этим он часто называется **парным коэффициентом корреляции Пирсона**. Коэффициент корреляции принято обозначать латинской буквой **r** Он рассчитывается по следующей формуле: **(СЛАЙД 11)**

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 * \sum_{i=1}^n (y_i - \bar{y})^2}}$$

где r_{xy} — коэффициент корреляции случайных величин X и Y ;

x_i, y_i — наблюдаемые значения величин X и Y для i -го объекта;

\bar{x}, \bar{y} — средние арифметические значения случайных величин;

n — количество наблюдаемых объектов.

Свойства коэффициента корреляции:

1. Коэффициент корреляции - величина безразмерная; значения ее заключаются в интервале $[-1, +1]$. **(СЛАЙД 12)**

2. Если $r_{xy} = 1$, то имеет место функциональная связь между величинами X и Y , если $r_{xy} = 0$ - то линейная связь между величинами отсутствует;

3. Если $r_{xy} > 0$, то имеет место положительная или прямая связь между величинами X и Y , если $r_{xy} < 0$, то связь отрицательная или обратная.

4. Сила корреляционной связи:

если $|r_{xy}| < 0.3$, то корреляция считается слабой,

если $0.3 < |r_{xy}| < 0.7$, то корреляция считается умеренной,

если $|r_{xy}| > 0.7$, то корреляция считается сильной.

Однако следует всегда помнить, что речь идет об оценке прямолинейной корреляционной связи, и что равенство нулю корреляционной связи – не означает отсутствие причинно-следственной связи, которая может иметь и нелинейную форму.

В силу того, что коэффициент корреляции обычно рассчитывают на основе выборочных данных, как и для других числовых характеристик, следует оценивать его статистическую значимость (достоверность)/

Оценка достоверности (статистической значимости) выборочного коэффициента корреляции (СЛАЙД 13) производится с помощью t -критерия Стьюдента, который определяют по формуле:

$$t = \frac{|r_{xy}|}{m_r}$$

где $|r_{xy}|$ — коэффициент корреляции случайных величин X и Y ;

m_r - средняя квадратическая ошибка коэффициента корреляции.

- Полученное расчётное значение t -критерия сравнивается с табличным критическим значением, при числе степеней свободы $n' = n - 2$ и выбранном уровне значимости P .
- Если расчётное значение t больше табличного критического значения $t_{0,05}$, ($t_{0,01}$, $t_{0,001}$), то делается вывод о наличии статистически значимой корреляционной связи при уровне значимости $P < 0,05$ ($< 0,01$; $< 0,001$).
- Если расчётное значение t меньше табличного критического значения $t_{0,05}$, то делается вывод об отсутствии статистически значимой корреляционной связи при уровне значимости $P > 0,05$.

Рассмотрим **пример** вычисления и оценки достоверности коэффициента корреляции (**СЛАЙД 14**) по данным эксперимента, в котором получены данные о толщине кожного рубца и времени, необходимом для его полного замораживания в целях криодеструкции (журнал “Хирургия”, 1985, № 7, с.129). Результаты расчета необходимых значений представлены в табл.1.

Таблица 1.

Расчет коэффициента корреляции между толщиной (X, мм.) кожного рубца и временем его замораживания (Y, мин.) при криодеструкции

№ наблюдения	x_i мм	y_i мин
1	3	0,6
2	5	1,0
3	8	1,6
4	9	1,5
5	12	1,7
6	14	1,8
7	17	2,4
8	20	3,0
$\sum_{i=1}^n$	88	13,6

$$\bar{x} = \frac{88}{8} = 11 \text{ мм}; \quad \bar{y} = \frac{13,6}{8} = 1,7 \text{ мин};$$

$$r_{xy} = 0,953; \quad t = 0,972 * \sqrt{\frac{8-2}{1-0,972^2}} = 10,13$$

При $n'=n-2=6$ критические значения t : $t_{05}=2,42$; $t_{01}=3,71$; $t_{001}=5,96$. Полученное $t > t_{001}$, следовательно, коэффициент корреляции статистически значим с весьма малой вероятностью ошибки — $P_0 < 0,001$.

Таким образом можно сделать вывод, что между экспозицией замораживания и толщиной кожного рубца имеется прямая ($r_{xy} > 0$), сильная ($r_{xy} > 0,7$) и статистически значимая ($P_0 < 0,001$) корреляционная связь.

Коэффициент корреляции рангов Спирмена

В случаях, когда результаты наблюдений x_i и y_i даются приближенно

или выражаются атрибутивными (качественными) оценками их интенсивности, например, в баллах, для приближенной оценки корреляционной связи применяют **коэффициент корреляции рангов Спирмена (СЛАЙД 15):**

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

где d_i - разность рангов величин x_i и y_i для i -го объекта наблюдения;

n – количество опытов (наблюдаемых объектов).

Ранги наблюдаемых значений X и Y являются числами натурального ряда 1, 2, ..., n , которые присваиваются им в порядке возрастания (или убывания) значений X и Y . В случае равенства отдельных значений X и Y для них вычисляется средний ранг.

Оценка достоверности коэффициента корреляции рангов производится по специальной таблице критических значений ρ или (при большем числе наблюдений) с помощью критерия t-Стьюдента. Разберё это на примере.

Пример 2. (тот же **СЛАЙД 15**) По данным Ю.С. Малеты и В.В. Тарасова ("Непараметрические методы статистического анализа в биологии и медицине", М., МГУ, 1982, с. 151) вычислить и оценить коэффициент корреляции рангов Спирмена между длительностью контакта рабочих с промышленными ядами и заболеваемостью токсическим гепатитом (см. табл.2).

Таблица 2.

Взаимосвязь между длительностью контакта с промышленными ядами (X , годы) и заболеваемостью токсическим гепатитом (Y , %)

Номер набл.	Стаж работы X , годы	Заболевае мость Y , %	Ранги		Разности d_i	d_i^2
			X	Y		
1	до 1	2	1	1	0	0

2	1 - 2	8	2	3	- 1	1
3	2 - 3	7	3	2	1	1
4	3 - 5	11	4	5	- 1	1
5	5 - 10	10	5	4	1	1
6	более 10	13	6	6	0	0
n = 6						$\sum d_i^2 = 4$

Решение (СЛАЙД 16) Коэффициент корреляции равен 0,89, что указывает на наличие между признаками прямой сильной корреляционной связи. Расчётное значение t равно 3,91. По таблице критических значений при $n' = 6 - 2 = 4$ $t_{0,05} = 2,78$, $t_{0,01} = 4,6$ корреляция статистически значима ($p < 0,05$).

Вывод: Между продолжительностью контакта с промышленными ядами и заболеваемостью токсическим гепатитом имеется прямая сильная корреляционная связь, коэффициент корреляции значим ($p < 0,05$).

Вопрос 3. Регрессионный анализ

Коэффициент корреляции характеризует направление и силу, или как говорят, тесноту связи между случайными величинами, но не отвечает на вопрос: какое значение может принять случайная величина Y , если другая, связанная с ней, случайная величина X примет какое-либо конкретное значение? Ответ на этот вопрос, имеющий важное значение для прогноза значений Y в зависимости от величины X , дает регрессионный анализ.

Термин «регрессия» впервые был введен в обиход математической статистики Френсисом Гамильтоном (1822-1911). Он исследовал проблемы наследственности, и в частности, физические и интеллектуальные способности ряда поколений нескольких сотен выдающихся семей. В ходе его исследования было выявлено, что дети в отношении своих индивидуальных показателей

стремятся воспроизвести средние значения показателей их родителей. Данное явление Гамильтон назвал «регрессией» (regression to mediocrity – возврат к среднему состоянию).

Каково назначение (цель) регрессионного анализа? (СЛАЙД 17 зачитать).

Итак, исходной предпосылкой является положение о том, что между величинами признака X и признака Y имеется зависимость, или как мы уже говорили, существует корреляционная связь. Такая зависимость в регрессионном анализе описывается тем или иным **уравнением**, связывающим среднее ожидаемое значение признака Y с возможными значениями признака x , которое называется **уравнением регрессии**.

Для простоты принимают, что между X и Y имеется линейная зависимость. Она описывается известным уравнением прямой линии:

$$Y = a + bx, \text{ (СЛАЙД 18)}$$

где a –е сть свободный член уравнения регрессии, а b носит название коэффициента регрессии. Их вычисление осуществляется по следующим формулам (СЛАЙД 19):

$$b = R_{y/x} = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$a = \bar{y} - b\bar{x}$$

Коэффициент регрессии $R_{y/x}$ (b) показывает на сколько единиц изменится параметр Y при изменении фактора X на одну единицу.

Свойства коэффициента регрессии (СЛАЙД 20):

1. Коэффициент регрессии - величина именованная с размерностью Y на X , например, при изучении регрессии массы на длину тела его размерность - кг/м, заболеваемости на температуру воздуха - %/с;

2. Величина коэффициента регрессии может быть целым, дробным, положительным или отрицательным числом;

3. Если $R_{y/x} = 0$, то при изменении x среднее значение y не изменяется;
 если $R_{y/x} > 0$, то при увеличении x величина y имеет тенденцию к увеличению;
 если $R_{y/x} < 0$, то при увеличении x величина y имеет тенденцию к уменьшению.

Поскольку коэффициенты уравнения регрессии получают по результатам ограниченного по объему выборочного исследования, они содержат ошибки репрезентативности, которые сказываются и на точности прогноза среднего ожидаемого значения \bar{y}_k . Величину ошибок прогноза оценивают средней квадратической ошибкой: **(СЛАЙД 21)**

$$m_{\hat{y}_k} = S_o \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

где S_o - среднее квадратическое отклонение наблюдавшихся значений величины y_i от рассчитанных по уравнению регрессии \hat{y}_i .

Эту величину вычисляют по формуле:

$$S_o = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}}$$

Из формул видно, что значение S_o и $m_{\hat{y}_k}$ будут тем меньше, чем меньше окажутся отклонения результатов наблюдения y_i от линии регрессии \hat{y}_i и чем больше будет число наблюдений n .

Для того, чтобы лучше понять и усвоить изложенный теоретический материал, предлагаю решить конкретную задачу.

По данным примера, представленного в таблице выше о криодеструкции кожного рубца, вычислим коэффициент и уравнение регрессии вида $\hat{y} = a + bx$;
 построим график линии регрессии;

сделаем прогноз времени экспозиции при толщине рубца $x_k = 15$ мм;

оценим точность и надежность прогноза (СЛАЙД 22).

Коэффициент регрессии $R_{y/x} = \frac{29,9}{240} = 0,125$ мин/мм.

Увеличение толщины кожного рубца на 1 мм требует увеличения экспозиции замораживания на 0,125 мин. или на 7,5 с.

Свободный член уравнения прямой $\hat{y} = a + R_{y/x}x$

$$a = 1,7 - 0,125 \cdot 11 = 0,329 \text{ мин};$$

уравнение регрессии y на x : $\hat{y}_k = 0,329 + 0,125 \cdot x_k$.

Прогноз экспозиции замораживания \hat{y}_k для заданной толщины кожного рубца $x_k = 15$ мм может быть выполнен а) по уравнению регрессии - $\hat{y}_k = 0,325 + 0,125 \cdot 15 = 2,2$ мин или 2 мин 12 с.;

Для построения графика линейной регрессии определяем по уравнению значения \hat{y}_k для крайних значения величины x (x_0 и x_{20}):

$$\hat{y}_0 = 0,325 + 0,125 \cdot 0 = 0,325;$$

$$\hat{y}_{20} = 0,325 + 0,125 \cdot 20 = 2,825.$$

Строим систему координат, откладывая по оси абсцисс значения величины x , а по оси ординат - величины y . Наносим на график точки, соответствующие значениям \hat{y}_0 и \hat{y}_{20} , и соединяем их прямой линией, которая и является линией регрессии (СЛАЙД 23, 24).

б) по графику линии регрессии - с таким же результатом. Оценка точности и надежности прогноза при

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0,215 \text{ мин}^2;$$

$$S_o = \sqrt{\frac{0,215}{8-2}} = 0,189 \text{ мин};$$

$$m_{\hat{y}_k} = 0,189 \sqrt{\frac{1}{8} + \frac{(15-11)^2}{240}} = 0,082 \text{ мин}.$$

95%-й доверительный интервал каждой величины \hat{y}_k определяется по формуле $\hat{y}_k \pm t_{0.05} m_{\hat{y}_k}$, где $t_{0.05} = 2,42$. Доверительный интервал прогноза $\hat{y}_{15} = 2,2 \pm 2,42 * 0,082 = 2,2 \pm 0,2$ мин.

Значения остальных интервалов, в пределах которых находятся истинные значения экспозиции замораживания, представлены в таблице и на рисунке (пунктирные линии).

Таблица 3

Необходимые расчеты для оценки точности надежности прогноза экспозиции замораживания кожного рубца в целях криодеструкции и в зависимости от его толщины

Толщина кожного рубца x_i мм	Экспозиция замораживания y , мин		Отклонения, мин $y_i - \hat{y}_i$	Квадраты отклонений $(y_i - \hat{y}_i)^2$	Отклонения $x_k - \bar{x}$ мм	Средние квадратичные ошибки $m_{\hat{y}_k}$	Доверительные интервалы $\hat{y}_k \pm t_{0.05} m_{\hat{y}_k}$ ($t = 2,42$)
	По данным наблюдения y_i	По уравнению регрессии y'_i					
3	0,6	0,700	-0,100	0,0100	-8	0,119	0,414±0,98
5	1,0	0,950	0,050	0,005	-6	0,096	0,718±1,18
8	1,6	1,325	0,275	0,0756	-3	0,076	1,125±1,52
9	1,5	1,450	0,050	0,0025	-2	0,071	1,278±1,62
12	1,7	1,825	-0,125	0,0156	1	0,068	1,661±1,98
14	1,8	2,075	-0,275	0,0756	3	0,076	1,875±2,27
17	2,4	2,450	-0,050	0,0025	6	0,096	2,218±2,68
20	3,0	2,825	-0,175	0,0306	9	0,128	2,515±3,13
$\bar{x} = 11$ мм	$\bar{y} = 1.7$ мм			$S(y_i - \hat{y}_j) = 0.215$			

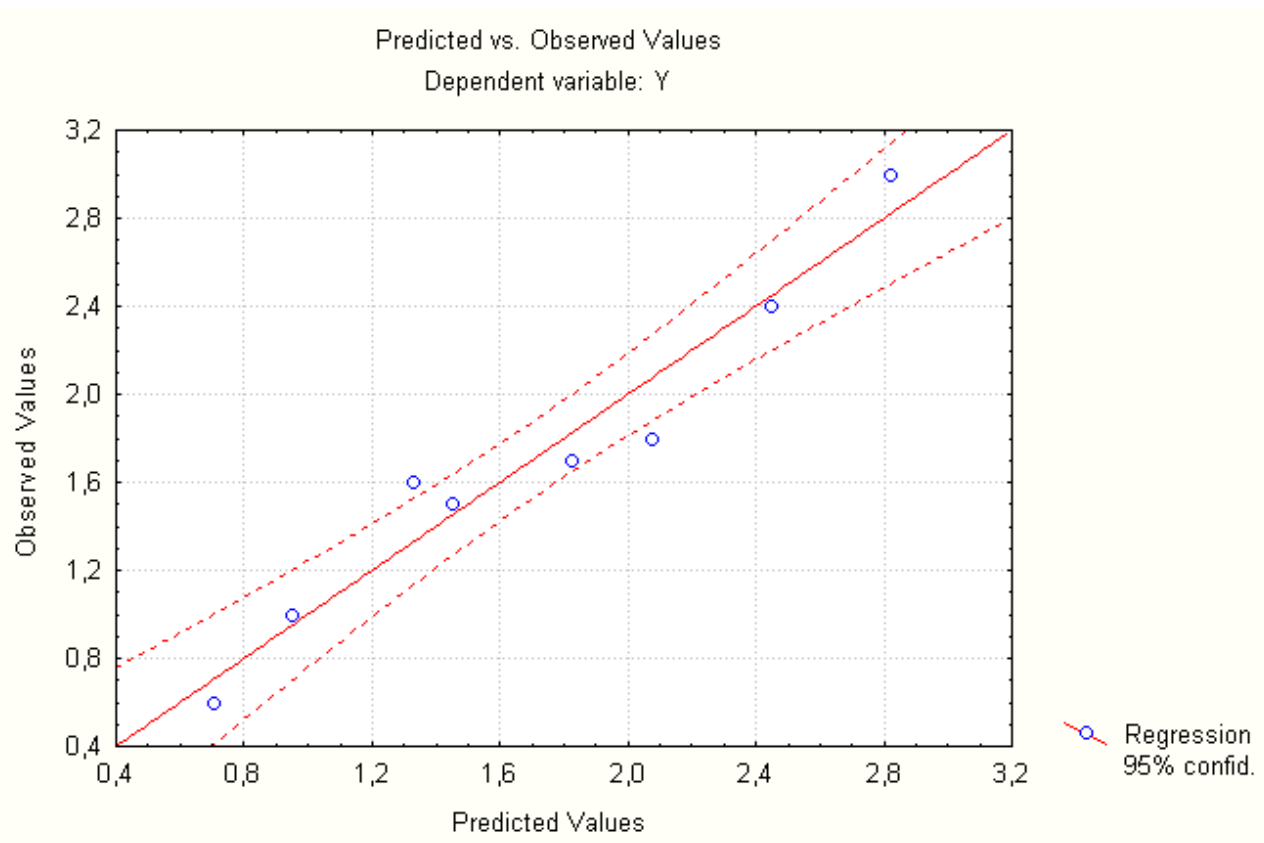


Рис.3. График линии регрессии.

Несколько упрощая, можно сказать, что целью регрессионного анализа является определение (прогноз) границ 95% доверительного интервала для математического ожидания параметра Y при любом заданном значении фактора X

Заключение.

Изучение методов связи между различными признаками должны найти широкое применение в научно-практической работе любого военного врача.

Использование корреляционного и регрессионного анализа позволяет количественно выразить зависимость между факторами-причинами и их следствиями, представить ее в виде математической (при необходимости графической) модели, осуществить необходимый прогноз. Освоение врачами корреляционно-регрессионного анализа будет способствовать объективизации результатов анализа деятельности подразделений, частей и учреждений медицинской службы, а в клинической медицине - клинко-диагностических заключений и тем самым повышению качества лечебной и лечебно-профилактической работы.

Существенно облегчит применение такого анализа, использование врачами современных мощных пакетов прикладных программ по статистической обработке данных на персональных компьютерах.

В лекции рассмотрены вопросы измерения связи только между парными количественными признаками при сравнительно малом числе наблюдений. Дальнейшее углубленное изучение корреляционно-регрессионного анализа должно кроме этого включать методы измерения линейной корреляции и регрессии при большом числе наблюдений, измерения связей между количественными и качественными признаками, измерения криволинейной связи, измерения множественной корреляции и регрессии. Необходимая литература по применению этих методик в медицине и биологии имеется.

Доцент

А.С. Кобзев

17 ноября 2022 г.