

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. Корнеев

« ____ » _____ 2022 г.

Преподаватель кандидат медицинских наук доцент А.Кобзев.

ЛЕКЦИЯ № 2

по дисциплине МЕДИЦИНСКАЯ СТАТИСТИКА И ИНФОРМАТИКА

на тему: «ОСНОВЫ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ»

для слушателей клинической ординатуры
клинического и медико-профилактического профиля

Обсуждена и одобрена на заседании кафедры
Протокол № 10
« 20 » октября 2022 г.

Уточнено (дополнено):

« ____ » _____ 20 ____ г. _____
(подпись)

СОДЕРЖАНИЕ

№ п/п	Учебные вопросы	Время (мин)
	Введение	5
1.	Понятие о статистических гипотезах и критериях значимости	15
2.	Методы оценки значимости различия относительных и средних величин	50
3.	Непараметрические критерии	15
	Заключение	5

Учебное время – 2 часа.

Литература

а) Используемая 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 сегодняшней лекции мы уже будем использовать те основные понятия и определения, с которыми вы познакомились на предыдущей лекции («Основы выборочного метода статистического исследования»). Напомню некоторые из них: сплошное и выборочное статистическое исследование, генеральная и выборочная совокупности, истинные и выборочные числовые характеристики, ошибки репрезентативности, доверительные вероятности.

Полагаю, что вы уже усвоили положение о том, что абсолютное большинство статистических исследований являются выборочными.

Обобщающие числовые характеристики, напомню вам, что это не что иное, как выборочные статистические показатели, полученные в ходе выборочных медико-статистических исследований, используются для анализа и объективной оценки результатов различных видов и направлений деятельности подразделений, частей и учреждений медицинской службы, а также тех или иных параметров здорового или больного организма. **(СЛАЙД 4) Важно знать, что анализ и оценка всегда носят сравнительный характер** и предполагают сравнение величин тех или иных изучаемых признаков с целью установления их изменений (во времени или при изменении тех или иных условий). Интересно то, что проанализировать единичный показатель не представляется возможным. Например, мы имеем информацию о том, что

средняя длительность лечения больных в терапевтическом отделении за год составила 21,2 дня. Или, частота послеоперационных осложнений в хирургическом отделении за год составила 4,6%. И что, это много или мало, плохо или хорошо? Оказывается, что для ответа на этот вопрос нам необходимо иметь некоторый аналогичный показатель для сравнения, который будет выполнять роль определенной контрольной величины. Как правило, принято рассчитывать статистические показатели за определенные периоды времени, в военно-лечебных учреждениях (организациях) чаще всего по результатам работы за год. При анализе принято сравнивать величину анализируемого показателя с аналогичным, полученным за предыдущий тот же период времени или с существующим нормативом. Вот тогда мы и сможем определить, в какую сторону произошло изменение показателя по величине и ответить на вопрос хорошо это или плохо. В специально проводимых статистических исследованиях, особенно клинических, достаточно часто сравнение показателей осуществляется в так называемых «опытных» и «контрольных» группах, а правильнее сказать выборках.

Другим важным моментом является то, что при анализе результатов выборочного медико-статистического исследования одной из наиболее частых и важных задач является задача количественной оценки значимости (достоверности) различий сравниваемых выборочных показателей (чаще всего средних и относительных величин).

В военной, да и гражданской медицине практически невозможно назвать ни одного исследования, в котором можно было бы обойтись без решения задачи ста-

статистической оценки значимости различий полученных обобщающих показателей (например таких, как оценка динамики показателей физического развития военнослужащих, сравнение эффективности лечения нескольких групп больных; оценка сдвигов в деятельности тех или других систем и органов человека или экспериментальных животных под воздействием различных экстремальных и других факторов; оценка основных и побочных действий лекарственных препаратов, оценка эффективности новых методов диагностики или методик лечения и т.п.). Любое клиническое, экспериментальное и лабораторное медицинское исследование, которое должно ответить **на вопрос об эффективности** применяемого метода лечения, не может быть признано научно обоснованным без статистической оценки достоверности изменений наиболее важных медицинских показателей. Так же строго следует подходить к решению аналогичных вопросов в военной эпидемиологии, военной гигиене, при организации медицинского обеспечения войск и сил флота, во всех случаях оценки эффективности и качества мероприятий, направленных на сохранение и улучшение здоровья всех групп и категорий военнослужащих как в мирное, так и в военное время.

1. Понятие о статистических гипотезах и критериях значимости

С точки зрения теории статистики эта задача укладывается в понятие *проверки статистической гипотезы*. (СЛАЙД 5)

Гипотеза вообще – это предположение, нуждающееся в доказательстве.

Под *статистической гипотезой* следует понимать всякое высказывание о

генеральной совокупности, проверяемое по выборке с применением специальных математико-статистических методов. Статистические гипотезы классифицируют на гипотезы о законах распределения и гипотезы о параметрах распределения. Так, например, гипотеза о том, что рост людей, имеет нормальный закон распределения, является гипотезой о законе распределения. Гипотеза о том, что средние значения артериального давления у больных гипертонической болезнью до и после приема гипотензивного препарата, различаются между собой, является гипотезой о параметрах распределения.

Проверка статистических гипотез выполняется на основе применения статистических критериев. **Статистический критерий** – это свод необходимых и достаточных правил, указывающих при каких результатах (условиях) проверяемая гипотеза принимается, а при каких отклоняется.

В медицинской статистике применяется большое число таких критериев, называемых критериями значимости или критериями согласия. Эти критерии могут быть *параметрическими*, когда в основе их применения лежит оценка числовых характеристик (параметров) совокупностей значений изучаемых признаков или законов (свойств) их распределения, и *непараметрическими*, не требующими знания этих параметров и законов. Большое число таких критериев носят имена авторов, разработавших теорию и практику их применения, и обозначаются, обычно, латинскими или греческими буквами (лямбда- критерий Колмогорова-Смирнова, Т- критерий Мани-Уитни, хи-квадрат критерий Пирсона, F-критерий, t- критерий Стьюдента, критерий Вилкоксона и др.).

Результатом применения статистического критерия является заключение о принятии или отклонении одной из двух конкурирующих гипотез, одна из которых, называемая **нулевой (H_0)**, предполагает отсутствие статистически значимых различий между сравниваемыми числовыми характеристиками, а вторая, противоположная ей, носит название **альтернативной гипотезы (H_1)** и предполагает наличие между этими характеристиками значимых различий.

Проверка гипотез осуществляется при определенной доверительной вероятности или уровне значимости, но не ниже 0,95 ($P < 0,05$). **Доверительная вероятность** – это вероятность правильности проверяемой гипотезы, а **уровень значимости** – это вероятность ошибки проверяемой гипотезы.

(СЛАЙД 6) (прошу сейчас особого внимания и сосредоточения в силу важности понимания далее излагаемого). Конструкция выдвигаемой и проверяемой статистической гипотезы должна совпадать с тем фактом или утверждением, которые необходимо доказать (проверить).

Поэтому проверяемая гипотеза формулируется следующим образом:

«Поскольку сравниваемые статистические показатели не равны между собой по величине, следует предположить, что между ними имеются статистически значимые (достоверные, существенные, значимые,) различия».

Сформулированная по указанному принципу гипотеза, подлежащая проверке, называется **рабочей**, она соответствует вышеназванной альтернативной гипотезе и обозначается как H_1 . Напоминаю, что противоположная гипотеза о том, что между сравниваемыми показателями достоверных различий нет, называется нуле-

вой и обозначается H_0 .

(СЛАЙД 7) При этом следует иметь в виду, что, принимая или отклоняя гипотезу H , можно допустить ошибки двух видов (табл. 1).

Таблица 1

Вероятностная оценка проверяемых гипотез

Гипотеза	Доверительная вероятность $1 - P$ (вероятность правильности альтернативной гипотезы)			
	$<0,95(95\%)$	0,95 (95%)	0,99 (99%)	0,999 (99,9%)
Альтернативная (H_1)	Отклоняется	Принимается с надежностью		
		95%	99%	99,9%
Нулевая (H_0)	Принимается	Отклоняется		
	$>0,05$	0,05%(5%)	0,01 (5%)	0,001(0,1%)
	Уровень значимости P (вероятность ошибки альтернативной гипотезы)			

Ошибка первого рода состоит в том, что нулевая гипотеза H_0 отвергается, т. е. принимается гипотеза H_1 , в то время как в действительности все же верна гипотеза H_0 . **Ошибка второго рода** состоит в том, что гипотеза H_0 принимается, в то время как верна гипотеза H_1 .

Вероятность ошибки первого рода принято обозначать через α , вероятность ошибки второго рода через β .

2. Оценка значимости различий выборочных относительных и средних величин

Проверка гипотезы о существенности (достоверности) различий двух выборочных статистических показателей может быть осуществлена с помощью разных

статистических процедур:

1-я из них позволяет это сделать предварительно, грубо и основывается на сопоставлении **доверительных интервалов**;

2-я позволяет сделать это точно и надежно с применением того или иного **статистического критерия значимости**.

Сопоставление доверительных интервалов для проверки гипотез производится, как правило, на низшем, 95%-ном уровне надежности и имеет, чаще всего, предварительное ориентировочное значение. Суть этой процедуры заключается в получении значений доверительного интервала для сравниваемых показателей и установлении степени их совпадения или различия.

Методика определения границ доверительного интервала была вам дана на предыдущей лекции: (**показатель \pm его максимальная ошибка репрезентативности.**)

$$I_p = p \pm t_{m_p}; \quad I_x = x \pm t_{m_x} \quad (n' = n - 1)$$

При сравнении полученных значений интервалов используются следующие правила:

1. Если доверительные интервалы сравниваемых показателей совпадают полностью или более чем на треть, то различие между показателями признается статистически не значимым, случайным, не существенным. Принимается нулевая гипотеза ($P > 0,05$), достоверных различий нет.

2. Если доверительные интервалы сравниваемых показателей совпадают менее чем на треть или вовсе не совпадают, то различие между показателями признается статистически значимым, неслучайным (достоверным, существенным) с веро-

ятностью не менее $\geq 95\%$ (уровнем значимости $P \leq 0,05$). Нулевая гипотеза отклоняется.

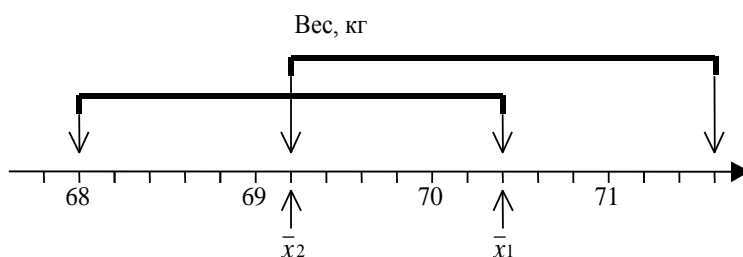
Разберем это на конкретном примере. **(СЛАЙД 8) Пример 1.** Изучались адаптационные сдвиги у 200 военнослужащих при воздействии повышенной температуры окружающей среды. Статистическая оценка различий некоторых из полученных авторами средних показателей представлена в таблице (вес, пульс, АД).

Таблица

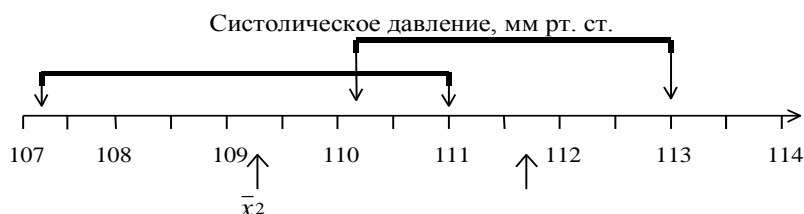
Оценка с помощью доверительных интервалов

Исследуемый физиологический показатель		\bar{x}	$m_{\bar{x}}$	$\bar{x} \pm 2m_{\bar{x}}$
Вес, кг	до воздействия	70,4	0,6	69,2÷71,6
	после воздействия	69,2	0,6	68,0÷70,4
Пульс, ударов/мин	до воздействия	67,3	0,6	65,3÷69,3
	после воздействия	76,2	0,7	73,9÷78,5
Систолическое давление, мм рт. ст.	до воздействия	111,6	0,7	110,2÷113,0
	после воздействия	109,2	0,9	107,4÷111,0

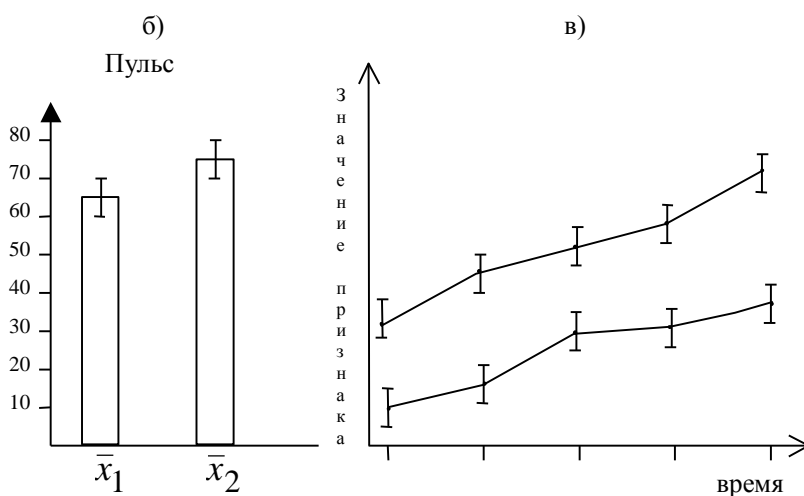
Как видно из данных таблицы, доверительные границы средних величин **веса военнослужащих** до и после воздействия повышенных температур (при доверительной вероятности 95%) совпали на протяжении половины каждого доверительного интервала. *Удобно сопоставлять доверительные интервалы графически.* Это выглядит так **(СЛАЙД 9)**. Следовательно, можно сделать вывод об отсутствии значимых различий между выборочными средними ($P_0 > 0,05$).



Доверительные границы средних величин *систолического давления* частично совпали, но это совпадение составило менее трети каждого из доверительных интервалов.



Имеются основания считать, что различие между этими средними является значимым ($P_0 < 0,05$). Доверительные границы средних значений **пульса** не совпали. Следовательно, различия между сравниваемыми средними статистически значимы ($P_0 < 0,05$).



Мы видим, что для улучшения наглядности при сравнении доверительных интервалов удобно использовать графические изображения, варианты которых представлены на слайде. Заметим, что вариант а) графического сопоставления интервалов проще для построения, относительно меньше по размерам и точнее для оценки различия. Вариант б), чаще применяемый в медицинской литературе, поз-

воляет наглядно сравнить не только доверительные интервалы, но и сами выборочные показатели по величине, однако он сложнее для построения и менее точен для оценки. Вариант в) применяют при сопоставлении статистических характеристик, изучаемых в динамике.

Применение статистических критериев для проверки гипотез

Как уже говорилось, более определенное и точное решение о принятии или отклонении проверяемой гипотезы позволяет сделать другой подход - **применение статистических критериев**. Методика применения большинства из критериев совпадает. Для ответа на вопрос о наличии или отсутствии значимых различий между сравниваемыми числовыми характеристиками необходимо: **(СЛАЙД 10)**

- Сформулировать рабочую гипотезу;
- Выбрать критерий;
- Получить расчетное значение критерия;
- Сравнить расчетное значение критерия с табличным критическим значением при выбранном уровне значимости;
- Сделать статистический вывод;
- Сделать логический вывод в терминах предметной области анализируемых данных.

Рассмотрим изложенную методику на примере применения известного и наиболее широко используемого (особенно в медицине) **параметрического критерия значимости t Стьюдента**, основой данного критерия является одноименное распределение, которое детально исследовал английский ученый Уильям Гос-

сет (псевдоним – «Стьюдент»).

Следует обратить внимание на то, что успешное применение критерия t возможно лишь в тех случаях, когда известно, что сравниваемый количественный признак распределен нормально или его распределение близко к нормальному. Если распределение признака существенно отличается от нормального или вовсе неизвестно, следует воспользоваться непараметрическими критериями оценки значимости различий.

Алгоритм, т.е. методика применения критерия t совпадает с рассмотренной ранее (СЛАЙДЫ 11 - 12).

1) определить по результатам выборочного исследования расчетное значение данного критерия, используя адекватную методику (формулу) его расчета;

2) определить, с учетом численности сравниваемых выборок (числа степеней свободы) и требуемой доверительной вероятности (уровня значимости), табличное (критическое) значение данного критерия;

3) сравнить величину двух полученных значений критериев по следующим правилам:

- если расчетное значение критерия не превышает (т.е. меньше или равно) табличного критического значения, следовательно, между сравниваемыми числовыми характеристиками статистически значимых различий нет, т.е. нулевая гипотеза H_0 принимается. ($P_0 > 0,05$);

- если расчетное значение критерия больше табличного критического значения, следовательно, между сравниваемыми числовыми характеристиками имеются

статистически значимые различия, т.е. нулевая гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1 ($P_0 < 0,05$ или $P_0 < 0,01$ или $P_0 < 0,001$ в зависимости от выбранного уровня значимости).

Существует несколько **формул расчета** критерия t , применение которых зависит как от вида сравниваемых числовых характеристик, так и от характеристик сравниваемых выборок. Общий подход к выбору адекватной формулы расчета величины критерия представлен на рис. (СЛАЙД 13)

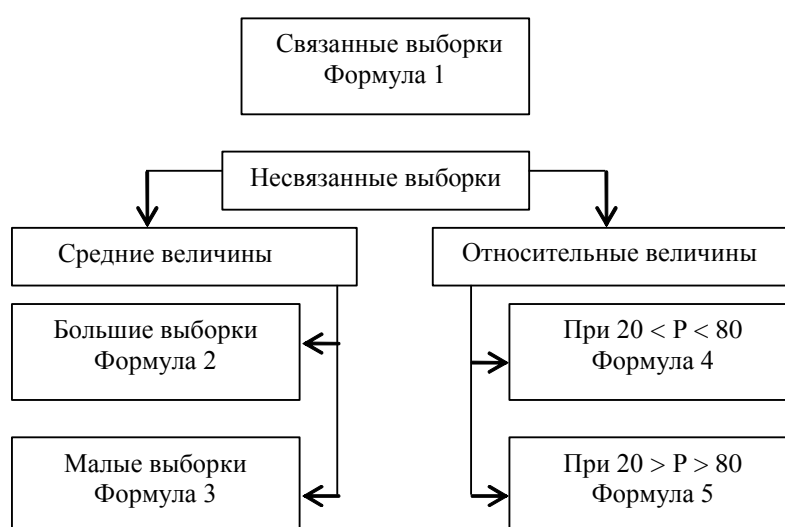


Рис. Методика выбора формулы расчета критерия t Стьюдента

При этом используются следующие расчетные формулы, которые приводятся в классическом и модифицированном вариантах, последний удобно применять при расчетах на вычислительной технике.

(СЛАЙД 14) **Формула 1.** Применяется в случае сравнения числовых характеристик, полученных на одной и той же группе объектов наблюдения, но в различных условиях, например, в начальный период заболевания и после лечения, или до воздействия некоторого фактора и после его воздействия. **Такие выборки, где результаты получены по принципу «до и после воздействия (опыта)», называ-**

ются связанными выборками.

Дано: результаты n пар наблюдений величин признака X до опыта – $x_1, x_2, \dots, x_i, \dots, x_n$; после опыта – $x'_1, x'_2, \dots, x'_3, x'_n$.

Величина критерия t для связанных выборок определяется по формуле:

$$t = \frac{\bar{\Delta}}{\sqrt{\frac{\sum (\Delta_i - \bar{\Delta})^2}{n(n-1)}}} = \frac{\sum (x'_i - x_i)}{\sqrt{\frac{n \sum (x'_i - x_i)^2 - [\sum (x'_i - x_i)]^2}{n-1}}},$$

где Δ_i – разность между величиной показателя после и до опыта ($\Delta_i = x'_i - x_i$); $\bar{\Delta}$ – среднее арифметическое значение разности, n – число парных наблюдений. Расчетная величина критерия t сравнивается с табличной при числе степеней свободы $n' = n - 1$. Приведенная методика в литературе носит также название расчета t критерия при сравнении совокупности с попарно связанными вариантами, или разностного метода.

Формула 2. (СЛАЙД 15) Применяется для оценки статистической значимости различия двух средних величин, полученных на независимых выборках при достаточно большой ($n > 30$) численности каждой из выборок:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{m_{\bar{x}_1}^2 + m_{\bar{x}_2}^2}},$$

где \bar{x}_1 и \bar{x}_2 – сравниваемые средние значения, а $m_{\bar{x}_1}$ и $m_{\bar{x}_2}$ соответствующие им средние квадратические ошибки; расчетная величина критерия t сравнивается с табличной при числе степеней свободы $n' = n_1 + n_2 - 2$.

Формула 3. (СЛАЙД 16) Применяется для оценки статистической значимости различия двух средних величин, полученных на независимых выборках при ма-

лом числе наблюдений ($n \leq 30$) в них, а также в случаях, когда число наблюдений в выборках сильно отличается:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{[S_1^2(n_1 - 1) + S_2^2(n_2 - 1)](n_1 + n_2)}{(n_1 + n_2 - 2)(n_1 * n_2)}}},$$

где \bar{x}_1 и \bar{x}_2 – сравниваемые средние значения, S_1 и S_2 – соответствующие им стандартные (среднеквадратические) отклонения, n_1 и n_2 – число наблюдений в каждой выборке; расчетная величина критерия t сравнивается с табличной при числе степеней свободы $n' = n_1 + n_2 - 2$.

Формула 4. (СЛАЙД 17) Применяется для оценки статистической значимости различий двух относительных величин (в %), полученных на независимых выборках при значениях показателей \bar{p}_1 и \bar{p}_2 не менее 20 и не более 80%:

$$t = \frac{|\bar{p}_1 - \bar{p}_2|}{\sqrt{m_{\bar{p}_1}^2 + m_{\bar{p}_2}^2}} = \frac{|\bar{p}_1 - \bar{p}_2|}{\sqrt{\frac{\bar{p}_1(100 - \bar{p}_1)}{n_1} + \frac{\bar{p}_2(100 - \bar{p}_2)}{n_2}}},$$

где \bar{p}_1 и \bar{p}_2 – сравниваемые значения показателей частоты (в %), $m_{\bar{p}_1}$ и $m_{\bar{p}_2}$ – соответствующие им средние ошибки, n_1 и n_2 число наблюдений в каждой выборке; расчетная величина критерия t сравнивается с табличной при числе степеней свободы $n' = n_1 + n_2 - 2$.

Формула 5. (СЛАЙД 18) Применяется для оценки статистической значимости различий двух относительных величин (в %), полученных на независимых выборках при значениях показателей \bar{p}_1 и \bar{p}_2 менее 20 и более 80%. В ней используется арксинусное преобразование относительных чисел, предложенное Р.А. Фишером

с целью обеспечить лучшее нормальное приближение при вычислении критерия t :

$$t = 2(| \arcsin \sqrt{\bar{p}_1/100} - \arcsin \sqrt{\bar{p}_2/100} |) * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}},$$

где \bar{p}_1 и \bar{p}_2 – сравниваемые значения показателей частоты (в %), n_1 и n_2 – число наблюдений в каждой выборке; расчетная величина критерия t сравнивается с табличной при числе степеней свободы $n' = n_1 + n_2 - 2$. Эта формула может быть с успехом использована при любых значениях сравниваемых относительных показателей.

Приведем **примеры** использования t критерия Стьюдента для определения значимости различий числовых характеристик с использованием рассмотренных формул.

Пример 2 (СЛАЙД 19). Имеются данные динамики содержания билирубина в крови 21 больного гепатитом А среднетяжелой формы (табл. 1). Требуется оценить значимость различия содержания билирубина в крови в различные периоды лечения больных. (**Комментарий** - Что приведено в таблице?)

Таблица 1

Содержание билирубина в крови (мкмоль/л) у больных гепатитом А среднетяжелой формы

№ больного	Содержание билирубина			№ больного	Содержание билирубина		
	при поступлении	на 10-й день	на 30-й день		при поступлении	на 10-й день	на 30-й день
1	60,3	87,0	8,5	12	104,2	152,6	19,6
2	71,8	98,4	12,6	13	121,8	142,7	14,8
3	105,4	81,4	18,6	14	98,4	168,5	20,7
4	100,4	141,5	31,6	15	64,3	148,5	25,2
5	110,8	136,5	12,4	16	48,4	131,5	15,4
6	90,4	126,7	9,6	17	75,6	148,5	18,4
7	132,6	169,4	30,2	18	121,8	169,5	18,4
8	128,0	168,4	81,4	19	103,7	151,8	15,6
9	98,6	138,4	10,2	20	158,6	161,3	26,8
10	130,0	171,0	20,6	21	138,1	158,6	13,6
11	128,0	170,0	18,6				

Вспоминаем алгоритм применения t-критерия Стьюдента

Решение. (СЛАЙД 20) Поскольку данные о содержании билирубина получены на одних и тех же единицах совокупности, выборки «При поступлении», «На 10-й день» и «На 30-й день» являются связанными. Следовательно, для получения расчетного значения критерия t необходимо воспользоваться формулой 1. Вычислим значение t при сравнении данных 1 и 2 выборки: $t_{\text{расч}} = 7,20$. Определим по таблице критическое значение t : при числе степеней свободы $n' = 21 - 1 = 20$ и уровне значимости $P = 0,05$, $t = 2,09$. Так как $t_{\text{расч}} > t_{0,05}$ можно сделать вывод, что содержание билирубина в крови больных среднетяжелой формой гепатита А на 10-й день болезни статистически значимо больше, чем при поступлении ($P_0 < 0,05$). Однако, можно установить, что расчетное значение t превышает и другие табличные значения: $t_{0,01} = 2,84$ и $t_{0,001} = 3,85$, следовательно вывод о наличии значимых различий можно сделать с существенно меньшим уровнем значимости ($P_0 < 0,001$), то есть с вероятностью 99,9%.

При сравнении содержания билирубина на 10-й и 30-й день болезни получим: $t_{\text{расч}} = 24,90$, а $t_{0,001} = 3,85$. Следовательно, с высокой надежностью ($P < 0,001$) можно говорить о существенном снижении содержания билирубина к 30-му дню лечения у этой категории больных.

Пример 3. (СЛАЙД 21) Исследовались данные физического развития рядового состава одной из воинских частей у солдат, проходящих службу по призыву и проходящих службу по контракту. Среднее значение веса военнослужащих 1-й группы составило $\bar{x}_1 = 69,4$ кг, при значении ошибки средней $m_{\bar{x}_1} = 1,4$ кг. Среднее

значение веса военнослужащих 2-й группы составило $\bar{x}_2 = 73,6$ кг, при значении ошибки средней $m_{\bar{x}_2} = 0,6$ кг. Требуется оценить существенность различий веса в исследуемых группах военнослужащих.

Поскольку выборки являются несвязанными, а численность их достаточно большой, для расчета критерия t необходимо воспользоваться 2-й формулой:

$$t = \frac{|x_1 - x_2|}{\sqrt{m_{\bar{x}_1}^2 + m_{\bar{x}_2}^2}} = \frac{|69,4 - 73,6|}{\sqrt{1,4^2 + 0,6^2}} = 2,76.$$

Табличное значение t возьмем при уровне значимости $P = 0,01$ и $n' \rightarrow \infty$; $t_{0,01} = 2,58$. Так как $t_{\text{расч}} > t_{0,01}$ можно сделать вывод о том, что среднее значение веса военнослужащих 2-й группы достоверно превышает среднее значение веса военнослужащих 1-й группы ($P < 0,01$).

Пример 4. (СЛАЙД 22) Исследовалась эффективность нового метода лечения больных по показателю длительности лечения. Полученные при этом числовые характеристики, приведены в таблице 2. Требуется определить значимость различия средней длительности лечения больных с заболеваниями А и Б при традиционном и новом методе лечения.

Таблица 2

Определение значимости различия средней длительности лечения

Исследуемые группы		Числовые характеристики				
		n	\bar{x}	S_x	m_x	t
Заболевание А	Традиционный метод	10	37,5	4,8	1,52	1,44
	Новый метод	12	34,6	4,2	1,33	
Заболевание Б	Традиционный метод	13	48,6	5,7	1,58	4,39
	Новый метод	15	39,8	4,9	1,27	

Поскольку сравниваются средние арифметические значения, а численность выборок мала, для расчета критерия t необходимо воспользоваться формулой 3.

Для группы больных с заболеванием А табличное значение критерия t определим при $n' = 10 + 12 - 2 = 20$ и уровне значимости $P = 0,05$; $t_{0,05} = 2,09$. Так как $t_{\text{расч}} < t_{0,05}$, можно сделать вывод, что при заболевании А новый метод лечения не приводит к достоверному снижению длительности лечения больных ($P_0 > 0,05$). При заболевании Б $t_{\text{расч}}(4,39) > t_{0,001}(3,71)$, следовательно средняя длительность лечения значимо уменьшается при применении нового метода лечения ($P_0 < 0,001$).

Пример 5. (СЛАЙД 23). Имеются результаты проверки эффективности нового антисептика при обработке им операционного поля. При 134 бактериологических исследованиях частота положительных высевок из смывов с операционного поля составила 23,1%. В 104 контрольных исследованиях (обработка операционного поля спиртом и эфиром) частота высевок составила 31,7%. На основании того, что высеваемость микрофлоры после применения нового антисептика была ниже на 8,6%, чем в контрольных исследованиях, авторы сделали вывод о его значительных преимуществах.

С помощью критерия значимости t проверим это заключение. Поскольку сравниваются относительные показатели частоты и их величина находится в интервале от 20 до 80%, критерий будем вычислять по формуле 4:

$$m_{\bar{p}_1} = \sqrt{\frac{\bar{p}_1(100 - \bar{p}_1)}{n_1}} = \sqrt{\frac{23,1(100 - 23,1)}{134}} = 3,6\%;$$

$$m_{\bar{p}_2} = \sqrt{\frac{\bar{p}_2(100 - \bar{p}_2)}{n_2}} = \sqrt{\frac{31,7(100 - 31,7)}{104}} = 4,6\%;$$

$$t = \frac{|\bar{p}_1 - \bar{p}_2|}{\sqrt{m_{\bar{p}_1}^2 + m_{\bar{p}_2}^2}} = \frac{|12,3 - 31,7|}{\sqrt{3,6^2 + 4,6^2}} = 1,47.$$

Поскольку табличное значение $t_{0,05}$ при $n' \rightarrow \infty$ и $1 - P = 0,05$ составит 1,96, следовательно $t_{\text{расч}} < t_{0,05}$, то есть между показателями частоты положительных высевов из смывов с операционного поля достоверных различий не выявлено ($P_0 > 0,05$). Заключение авторов не верно.

Пример 7. (СЛАЙД 21) С целью профилактики гриппа (вирус В) изучалась эффективность интраназального введения жидкой вакцины. Из 280 человек, получавших вакцину, заболело 30 ($p_1 = 10,7\%$). В контрольной группе (не получавших вакцину) из 279 человек заболели 58 ($p_2 = 20,8\%$). Определить достоверность различий показателя частоты заболеваемости гриппом в обеих группах.

Решение. Поскольку один из сравниваемых показателей частоты имеет малое значение ($p_1 < 20\%$), для расчета критерия t необходимо воспользоваться формулой 5:

$$t = 2(\arcsin \sqrt{\bar{p}_1/100} - \arcsin \sqrt{\bar{p}_2/100}) * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}} =$$

$$t = 2(\arcsin \sqrt{10,7/100} - \arcsin \sqrt{20,8/100}) * \sqrt{\frac{280 * 279}{280 + 279}} = 3,32$$

Так как $t_{\text{расч}} > t_{0,001}$ (при $n' \rightarrow \infty$ и $1 - P = 0,001$, $t_{0,001} = 3,29$), следовательно, частота заболеваемости гриппом в группе привитых статистически значимо меньше, чем в группе не привитых ($P_0 < 0,001$).

3. Непараметрические критерии для проверки статистических гипотез

Наряду со статистическими методами анализа результатов клинических, экспериментальных и других исследований, основанных на оценке достоверности их выборочных характеристик (параметров), в последние годы получили распространение так называемые непараметрические методы статистической обработки наблюдений. Применение этих методов не требует знания закона распределения изучаемых совокупностей и вычисления их основных статистических характеристик (средних значений, среднеквадратических отклонений и ошибок, частот и т.п.). Их также целесообразно применять при малом числе наблюдений (до 30). Кроме того, достоинством ряда непараметрических методов является простота их использования, а также возможность приложения к результатам выборочного наблюдения, имеющим полуколичественное выражение (например, оценка тяжести состояния больного, результатов его лечения: полное выздоровление, улучшение, без изменения, ухудшение, смерть; оценка интенсивности биологической реакции: —, ±, +, ++, +++ , +++++ и т.д.).

Критерий согласия χ^2 Пирсона

Во многих практических военно-медицинских задачах точный закон распределения исследуемой случайной величины, неизвестен, т. е. является гипотезой, которая требует статистической проверки. Для этого следует проверить нулевую гипотезу (H_0) о том, что эта случайная величина подчиняется выбранному теоретическому закону распределения $F(x)$. С этой целью обычно производят выборку, состоящую из n независимых наблюдений над случайной величиной X . Далее по выборке строят эмпирическое распределение $F^*(x)$ исследуемой случайной величины. Сравнение эмпирического $F^*(x)$ и теоретического распределений производится с помощью специально подобранной статистики — критерия согласия. Существует несколько критериев согласия: χ^2 Пирсона, Колмогорова, Смирнова и др.

Критерий согласия χ^2 Пирсона («хи»-квадрат) — один из первых разработанных и наиболее часто употребляемый критерий для проверки гипотезы о законе распределения или сравнения двух эмпирических распределений. Критерий Пирсона χ^2 рассчитывается по формуле:

$$\chi^2 = \sum_{i=1}^m \frac{(f - f_m)^2}{f_m}$$

где f — эмпирические частоты, а f_m — теоретические частоты.

При числе степеней свободы $n' = l - r - 1$, где l — число групп (интервалов) в совокупности, r — количество параметров (общих характеристик) теоретического распределения.

Данный критерий может быть использован и для сравнения эмпирических законов распределения. В этом случае он рассчитывается по следующей формуле:

$$\chi^2 = \sum_{i=1}^m \frac{(p_{1i} - p_{2i})^2}{p_{2i}}$$

где p_{1i} — наблюдаемые частоты в первой выборке, а p_{2i} — наблюдаемые частоты во второй выборке. При числе степеней свободы $n' = (S-1)(r-1)$, где S — число граф (столбцов), а r — число строк в таблице (без итогов).

Правило применения критерия χ^2 сводится к следующему. Рассчитав значение χ^2 и выбрав уровень значимости критерия α при числе степеней свободы n' , по таблице χ^2 -распределения (Приложение 5) определяют $\chi^2_{r;\alpha}$. Если $\chi^2 > \chi^2_{r;\alpha}$, то гипотезу H_0 отвергают, если $\chi^2 \leq \chi^2_{r;\alpha}$, то гипотезу принимают. Очевидно, что при проверке гипотезы о законе распределения контролируется лишь ошибка первого рода.

Необходимым условием применения критерия Пирсона χ^2 является наличие в каждом из интервалов по меньшей мере 5—10 наблюдений. Если количество наблюдений в отдельных интервалах очень мало (порядка 1—2), то имеет смысл объединить некоторые интервалы.

Пример 10: По данным осеннего обследования 500 военнослужащих определить подчиняется ли их рост закону нормального распределения с помощью χ^2 . Результаты решения представлены в таблице 4.6.

Таблица 4.6

Данные роста 500 военнослужащих по данным осеннего обследования

Интервал роста во- еннослу- жащих	Число военно- служа- щих	Сере- дина интер- вала	$x - \bar{x}$	Норми- рован- ное значе-	Плот- ность рас- пре-	Теоретиче- ская частота	$f - f_m$	$\frac{(f - f_m)^2}{f_m}$
--------------------------------------------	-----------------------------------	---------------------------------	---------------	-----------------------------------	--------------------------------	----------------------------	-----------	---------------------------

	f	x		ние ро- ста $t = \frac{x - \bar{x}}{\sigma}$	деле- ния $\varphi(t)$	$f_m = \frac{h \sum f}{\sigma}$ $\times \varphi(t)$		
150-155	4	152.5	-20.35	-2.76	0.0088	3	1	0.33
155-160	10	157.5	-15.35	-2.09	0.0449	15	-5	1.67
160-165	61	162.5	-10.35	-1.41	0.1476	50	11	0.22
165-170	100	167.5	-5.35	-0.73	0.3056	104	-4	0.15
170-175	130	172.5	0.35	-0.05	0.3984	136	-6	0.26
175-180	114	177.5	4.65	0.63	0.3271	11	3	0.08
180-185	62	182.5	9.65	1.31	0.1691	58	4	0.27
185-190	11	187.5	14.65	1.99	0.0551	19	-8	3.37
190-195	8	192.5	19.65	2.67	0.0113	4	4	4.00
	500					500		10.35

$$\chi^2 = \sum_{i=1}^m \frac{(f - f_m)^2}{f_m} = 10,35$$

Приняв уровень значимости $\alpha=0,05$, находим, что для $n'=9-2-1=6$ $\chi^2_{\alpha \alpha \alpha \alpha} = 12,59$. Так как фактическое $\chi^2 < \chi^2_{\alpha \alpha \alpha \alpha}$, то отклонения фактических частот от теоретических можно считать случайными, а само распределение по росту 500 военнослужащих — близким к нормальному.

Z-критерий знаков

Данный критерий является наиболее простым непараметрическим критерием для оценки статистической значимости различия двух связанных выборок. Особенностью этого критерия является то, что при его расчете не учитываются абсолютные величины различий сравниваемых парных наблюдений (например, результатов каких-то лабораторных показателей у одних и тех же больных, до и после лечения), а только их направленность (уменьшение или увеличение). Эта направленность обычно обозначается знаками «+» или «—» (откуда и произошло название критерия).

Методика расчета критерия знаков.

Из первого значения производят вычитание второго значения для каждой пары наблюдений. Полученные результаты обозначают знаками «+» если результат больше 0, и «—» - в противоположном варианте. Не имеющие изменений пары, из

дальнейшей оценки исключаются.

Подсчитывается общее число (n) парных наблюдений, имеющих различия (т.е. отмеченных знаками «+» и «—»).

Подсчитывается число знаков, которое реже всего встречается.

Полученное число (z) сравнивается (с учетом n) с критическими значениями z_0 , содержащимися в Приложении 6. Если найденная величина z больше или равна табличному z_0 , принимается нулевая гипотеза. Если $z < z_0$, то различия между сравниваемыми связанными совокупностями могут считаться значимыми с соответствующими уровнями вероятности ($P < 0,05$ или $P < 0,01$).

Пример 11.

У 12 работающих на ультразвуковых установках изучалось содержание сахара в крови натощак до работы и через три часа после работы. Исходные данные в таблице 4.7.

Таблица 4.7.

Содержание сахара в крови обследованных натощак до работы
и после 3 часов работы на ультразвуковых установках

пп	С АХ_Д О	С АХ_П ОС	пп	С АХ_Д О	С АХ_П ОС
	1	5		6	6
12	4		4	6	
	8	6		7	6
2	7		0	6	
	1	9		8	4
01	6		8	8	
	7	5		8	5
2	9		0	1	0
	7	7		6	6
9	9		1	6	1

	8	7		8	6
2		6	2	8	1

Решение выполним с помощью непараметрического критерия знаков (Sign test). Результаты решения в машинограмме 4.1.

Машинограмма 4/1.

Sign Test (pr_1_4_4.sta)

	N	P		
	o. of	ercent		
	N	v	Z	P
	on-ties	< V		-level
CAX_ДО	1	9	2,4	0
& CAX_ПОС	1	,0909	12091	,0158
				61

Анализ решения. Снижение уровня содержания сахара в крови через 3 часа работы на ультразвуковых установках по сравнению с его уровнем натощак существенное с уровнем значимости $p=0,016$, а достоверность различия $1-p=1-0,016=0,984$ или 98,4%.

Парный критерий Вилкоксона

Критерий учитывает не только направленность, но и величину разности парных вариантов. Это обуславливает его большую статистическую мощность по сравнению с критерием знаков. Данный критерий является аналогом t-критерия Стьюдента для связанных выборок.

Методика расчета парного критерия Вилкоксона:

Найти разности сравниваемых парных наблюдений.

1. Определить порядковые номера (ранги) абсолютных значений (т. е. без учета алгебраических знаков) полученных разностей в порядке их возрастания (варианты с нулевыми разностями из дальнейшей оценки исключаются, а n — число пар — соответственно уменьшается).
2. Подсчитать суммы рангов для разностей, имеющих одинаковые алгебраические знаки, и взять меньшую из этих сумм (T^{Δ}).

Далее следует сравнить эту сумму (T^Δ) с критическими значениями $T_{\varepsilon\delta\varepsilon\delta}^\Delta$ из Приложения 7 с учетом величины n . Если найденная величина $T^\Delta \geq T_{\varepsilon\delta\varepsilon\delta}^\Delta$, принимается нулевая гипотеза. Если $T^\Delta < T_{\varepsilon\delta\varepsilon\delta}^\Delta$, то различия между сравниваемыми связанными выборками признаются статистически значимыми.

Пример 12.

Для сравнения двух методов определения времени свертываемости крови, каждая проба оценивается этими двумя методами:

по Бюркеру - появление нитей фибрина при комнатной температуре;

по Ли-Уайту - при опрокидывании пробирки в термостате при температуре 37 градусов Цельсия кровь не выливается.

Исходные данные в таблице 4.8.

Таблица 4.8.

Время свертывания крови при его определении двумя методами

№ пп	BURKE R	LIWITE	№ пп	BURKE R	LIWITE
1	10	10	7	5	6
2	9	8	8	5	6
3	8	9	9	6	7
4	8	10	10	6	7
5	7	6	11	7	9
6	7	10			

Решение задачи осуществлено с помощью критерия Вилкоксона, который является непараметрической альтернативой t-критерию Стьюдента для парных сравнений количественных данных в зависимых выборках. Результаты решения в машинограмме 4.2.

Машинограмма 4/2.

Wilcoxon Matched Pairs Test (t_w.css)

	Valid N	T	Z	p-level
BURKER & LIWITE	11	8	1,988	0,047

Анализ результатов решения. Уровень значимости различия исследуемого

показателя $p = 0,047$, а достоверность его различия $1-p = 1-0,047 = 0,953$ или 95,3%, что свидетельствует о значимом различии времени свертывания крови при использовании исследуемых методов. Можно предположить, что механизм методов связан с различными звеньями процесса свертывания крови.

ЗАКЛЮЧЕНИЕ

Ещё раз напоминаю, поскольку в реальной практике работы подразделений, частей и учреждений медицинской службы собираемая и анализируемая информация всегда представляет собой лишь выборку из некоторой генеральной совокупности, расчет тех или иных обобщающих показателей должен проводиться с учетом требований выборочного метода исследования и сопровождаться оценкой их точности и надежности. В последующих лекциях, где будет продолжено изложение методов статистической обработки данных, а также на практических занятиях, при решении тех или иных задач, вы убедитесь, что правильный анализ информации и формулирование статистических выводов могут быть выполнены только с определенной долей вероятности, что обусловлено, в первую очередь, выборочным характером собираемых и анализируемых данных.

Доцент

А.Кобзев

« 18 » октября 2022 г.