Институт лидерства и управления здравоохранением
Сеченовский Университет
наук о жизни

Инструменты статистического анализа научных данных
Аналитическая статистика
Проверка статистических гипотез

Косаговская Ирина Игоревна

Высшая школа управления здравоохранением

23.11.2023

В медицине и здравоохранении часто используются, сознательно или неосознанно, различные статистические концепции при принятии решений по таким вопросам как

- оценка состояния здоровья его прогноз;

- выбор стратегии и тактики профилактики и лечения;

- оценка отдаленных результатов и выживаемости.

Статистика – наука, изучающая количественные закономерности материальных явлений в неразрывной связи с их качественной стороной.

Биостатистика - приложение общей теории статистики для решения научно-практических проблем в области биологии, медицины и здравоохранения.

Медицинская (отраслевая) статистика:

- **Статистика здоровья населения**
- **Статистика системы здравоохранения**

Статистика –

- это **инструмент** для анализа экспериментальных данных и результатов популяционных исследований;

- это **язык**, с помощью которого исследователь сообщает полученные им результаты и благодаря которому он понимает медико-статистическую информацию;

- это **элемент доказательной медицины**;

- это **база** для обоснования принятия управленческих **решений**.

Концепция «медицины, основанной на доказательствах» (evidence-based medicine, EBM) предполагает, что к широкому применению в медицинской практике должны рекомендоваться те вмешательства, эффективность и безопасность которых подтверждена в высококачественных исследованиях, выполненных на основе единых методологических принципов.

Вероятностная природа медицины и вероятностный характер процессов в общественном здоровье позволяют повсеместно применять методы математической статистики и теории вероятностей и выбирать их в зависимости от уровня решаемых задач с целью сведения к минимуму набора и степени проявления случайностей.

Основным инструментом для этого является теория вероятностей – математическая наука, устанавливающая закономерности случайных явлений.

Вероятность – количественная мера объективной возможности появления события при реализации определенного комплекса условий.

Вероятность события А обозначается как р(А) и выражается в долях единицы или в процентах.

Мера вероятности – диапазон ее числовых значений: от 0 до 1 или от 0 до 100%.

Случайное событие – событие, которое при реализации определенного комплекса условий может произойти или не произойти.
Его вероятность будет находиться в пределах
$0 < p(A) < 1$ или $0 < p(A) < 100\%$.

Достоверное событие - событие, которое при реализации определенного комплекса условий произойдет непременно.
Его вероятность будет равна **1 или 100%.**

Невозможное событие - событие, которое при реализации определенного комплекса условий не произойдет никогда.
Его вероятность будет **равна 0.**

В медицинских исследованиях достаточной считается вероятность появления события не менее 0,95 или 95%.

При изучении заболеваний или ситуаций, имеющих важнейшие медико-социальные последствия или высокие показатели летальности и инвалидности, а также при фармакологических исследованиях вероятность появления события должна быть не менее 0,99 (99%).

Частота появления события (статистическая вероятность) – это отношение числа случаев, в которых реализовался определенный комплекс условий (m), к общему числу случаев (n):

$$p(A)=m/n.$$

Вероятность отсутствия события: $q= 1- p$.

Например, если вероятность того, что больной останется жив спустя пять лет после операции равна 0.8 (80%), то вероятность того, что он за этот временной промежуток умрёт равна 0.2 (20%).

Этапы научно-практического исследования

1. Формулирование цели и задач исследования.
2. Организация исследования.
3. Сбор информации.
4. Обработка информации.
5. Анализ результатов исследования.
6. Внедрение результатов исследования в практику и оценка эффективности внедрения

Неверно организованный эксперимент не спасет никакой статистический анализ!

II этап: Организация исследования

В зависимости от степени охвата объекта исследования принято различать

- сплошное исследование

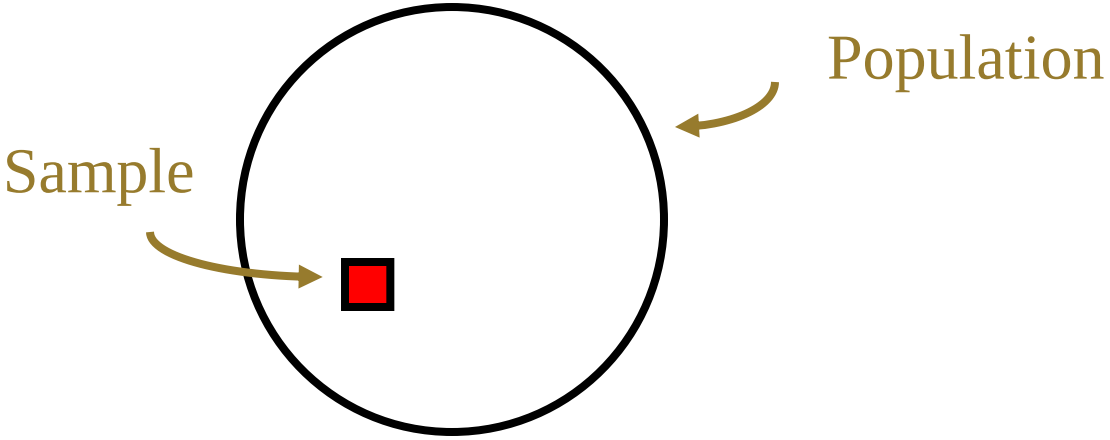
- выборочное исследование

II этап: Организация исследования

- **Сплошным** называется такое наблюдение, при котором изучаются все единицы наблюдения объекта исследования, т.е. так называемая **генеральная** совокупность **(population).**

II этап: Организация исследования

- **Выборочное наблюдение** - это вид несплошного наблюдения, при котором отбор подлежащих обследованию единиц наблюдения осуществляется случайно из генеральной совокупности, после чего результаты распространяются на всю исходную совокупность.
- Сформированная таким образом совокупность называется выборочной или **выборкой (sample).**



II этап: Организация исследования

- **Репрезентативность** - это представительность выборочной совокупности по отношению ко всей (генеральной)совокупности
- **Репрезентативность** должна быть **количественной и качественной.**

II этап: Организация исследования

- Под **количественной репрезентативностью** понимают **достаточное число единиц наблюдения в выборке** для проявления закона больших чисел.

- Под **качественной репрезентативностью** понимают соответствие признаков у единиц наблюдения генеральной и выборочной совокупностей.

Закон больших чисел:

при достаточно большом числе наблюдений случайные отклонения взаимно погашаются и проявляется основная тенденция (закономерность).

Расчет объема выборки

(определение необходимого числа наблюдений для исследования)

надо выполнить на этапе организации исследования!

II этап: Организация исследования

- **Репрезентативность выборки зависит**

➢ **от её численности и**

➢ **от способов формирования выборочной совокупности**, т.е. способов отбора единиц наблюдения (методов рандомизации).

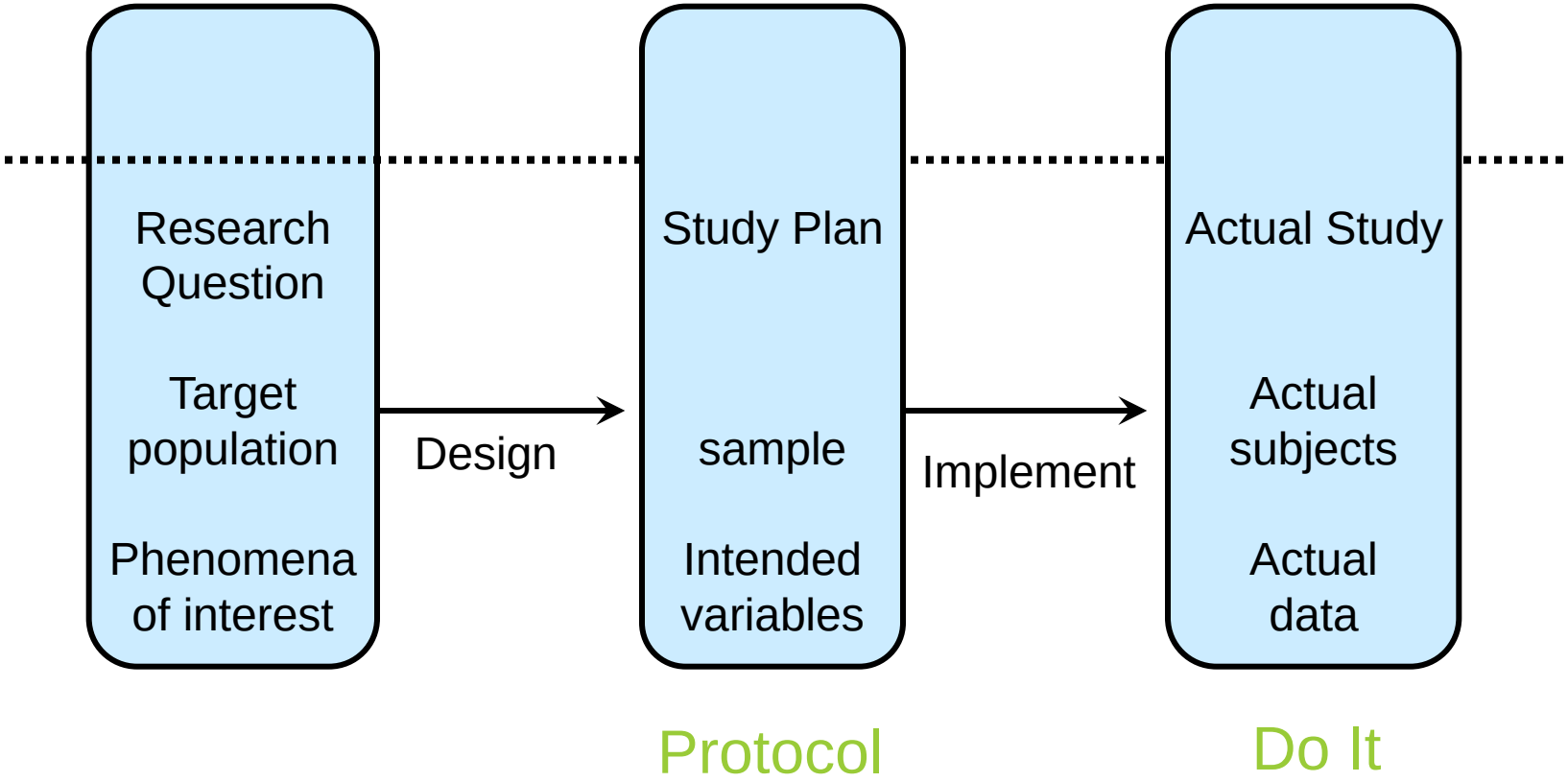
- Главное требование, предъявляемой к отбору - это его **случайность (рандомизированный отбор).** При этом каждой единице наблюдения обеспечивается одинаковая вероятность попадания в выборку благодаря случайности отбора.
- Случайность отбора достигается путем выбора и применения адекватного **метода рандомизации**, от которого будет зависеть полноценность получаемых данных и ,в конечном итоге, успех всего исследования.

II этап: Организация исследования

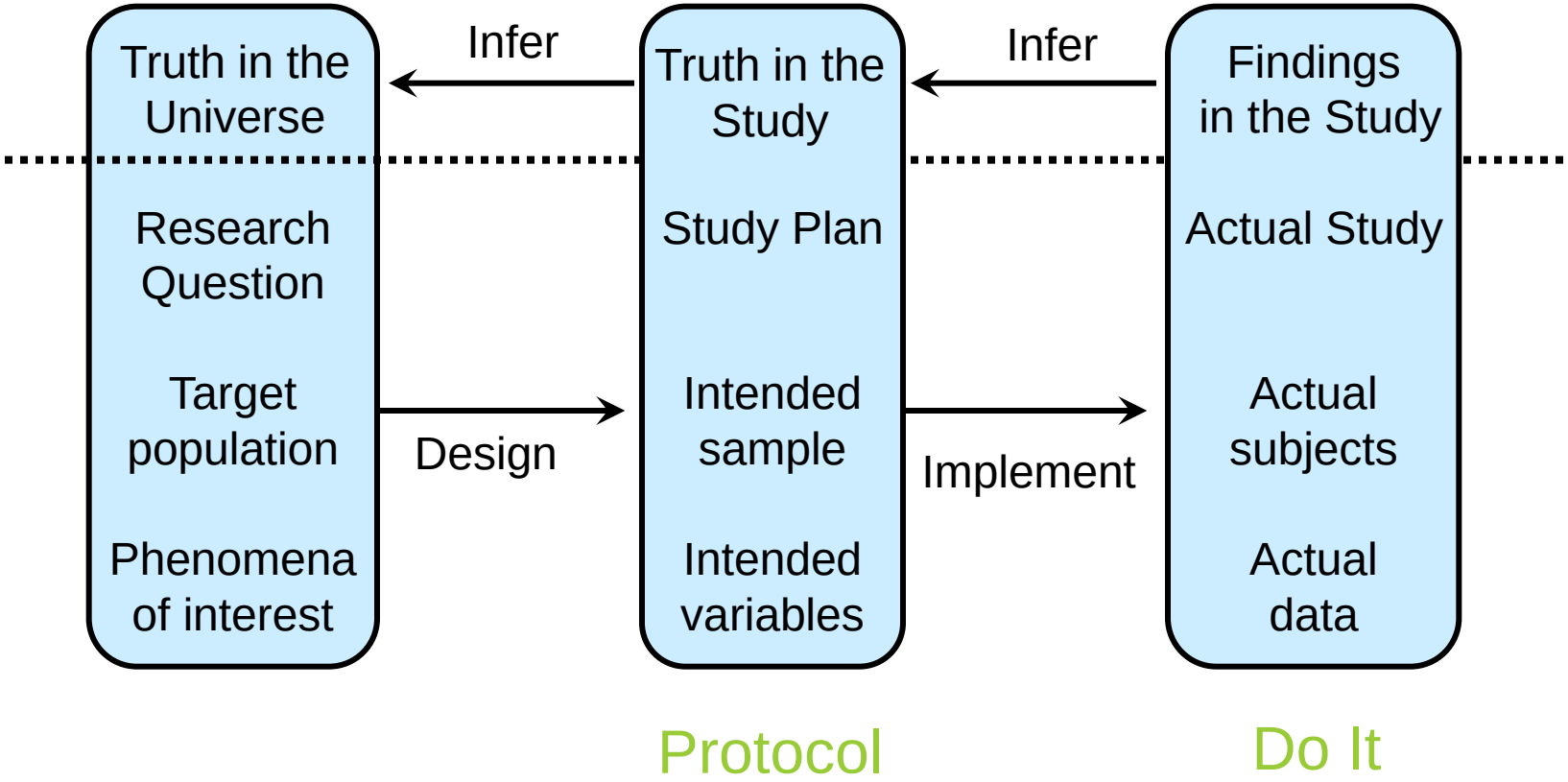
Типы структуры (дизайн) исследований в порядке возрастания доказательности:

1. **Описание отдельных случаев**
2. **Описание серии случаев (case series)**
3. **Исследование случай-контроль (case-control study)**
4. **Одномоментное исследование (cross-sectional study)**
5. **Когортное исследование (cohort study, follow-up)**
6. **Рандомизированное исследование (randomised study)**
7. **Мета-анализ (meta-analysis)**

Sample vs Target Population



Sample vs Target Population



ОЦЕНКА РЕЗУЛЬТАТОВ ВЫБОРОЧНОГО МЕТОДА ИССЛЕДОВАНИЯ

1 этап - точечная оценка параметра

2 этап – интервальная оценка параметра

Точечная оценка параметра выражается в **ошибке репрезентативности (standard error - SE, m)**, которая показывает на сколько отличаются обобщающие коэффициенты (показатели), полученные при выборочном исследовании, от тех коэффициентов, которые могли бы быть получены при сплошном исследовании.

Для количественных признаков:

$$m = \frac{\delta}{\sqrt{n-1}}$$

Для номинальных и порядковых признаков:

$$m = \pm\sqrt{\frac{pq}{n-1}}$$

Центральная предельная теорема

Для бесконечного числа независимых случайных выборок одинакового объема, извлеченных из генеральной совокупности, выборочное распределение любой линейной комбинации выборочных средних будет стремиться к нормальному при объеме выборки, стремящейся к бесконечности.

Интервальная оценка параметра

На практике определяют пределы возможных ошибок выборки или **предельную ошибку выборки (Δ).**

$$\Delta = tm$$

Т.к. предельная ошибка может быть как в сторону увеличения, так и в сторону уменьшения, то говорят о **доверительном интервале (ДИ)** или **доверительных границах (confidence interval, CD),** в пределах которых будет находиться показатель генеральной совокупности на основании данных выборочного исследования.

$$X_{выб} \pm \Delta;\ X_{выб} \pm tm;$$

$$X_{выб} - tm < X_{ген} < X_{выб} + tm$$

Существует таблица соответствия интеграла вероятности и значения “t” (таблица Стьюдента).

- При числе наблюдений больше 30 можно пользоваться следующей закономерностью:

вероятности безошибочного прогноза $p = 0{,}68$ приблизительно соответствует $t \cong 1$;

при $p = 0{,}95$ $t \cong 2$,

при $p = 0{,}99$ $t \cong 3$.

Доверительный интервал

- Диапазон значений, построенный по выборке, который с определенной степенью вероятности содержит истинное значение числового параметра генеральной совокупности.
- Это мера точности оцениваемого параметра.

- В диапазоне удвоенной стандартной ошибки по обе стороны от среднего значения ($\bar{X} \pm 2m$) с вероятностью примерно 95 % находится среднее значение генеральной совокупности.
- С вероятностью примерно 99 % оно лежит в диапазоне утроенной стандартной ошибки ($\bar{X} \pm 3m$).

- Для большинства медицинских исследований допускают р = 0,95 или 95%. В этом случае вероятность выхода результата выборочного исследования за границы доверительного интервала, т.е. вероятность ошибки составляет 0,05 или 5%.Поэтому говорят, что результат исследования получен с **уровнем значимости** 0,05 (р=0,05).

- При необходимости более строгой оценки р=0,99 (99%), вероятность ошибки составит 0,01 (1%) и следовательно **уровень значимости** будет р=0,01

Пример

В результате проведенного маммографического обследования 2000 женщин старше 35 лет, проживающих в одном из районов г.Н, у 20,0% из них были выявлены предраковые состояния молочной железы.

Генеральной совокупностью в данном случае являются все женщины старше 35 лет, проживающие в данном районе г.Н.

Точечная оценка параметра (расчет ошибки репрезентативности для номинальных и порядковых признаков):

$$m_p = \pm\sqrt{\frac{pq}{n}} = \pm\sqrt{\frac{20\%*80\%}{2000}} = \pm\sqrt{\frac{1600\%}{2000}} = \pm\sqrt{0.8\%} = \pm 0.89\ \%$$

где p - частота признака в изучаемой совокупности (20%)

q - альтернативный "p" показатель (100%-20%=80%)

n – число наблюдений (2000)

Интервальная оценка параметра (определение доверительного интервала - ДИ)

Будем определять доверительный интервал с вероятностью безошибочного прогноза 0,99 (99%) или вероятностью ошибки (уровнем значимости) 0,01 (1%). Тогда доверительный критерий «t» при условии, что число наблюдений больше 30 (в нашем случае – 2000) можно условно принять равным 3.

Теперь найдем доверительный интервал:

$P_{ген}=P_{выб} \pm tm = 20\% \pm 3*0.89\%$ 17.33 % $\leq p \leq$ 22.67 %

Вывод: **По данным выборочного обследования, установлено, что распространенность предраковых состояний молочных желез у женщин старше 35 лет, проживающих в данном районе г.Н, находятся в пределах от 17,3% до 22,7% (p=0,01)**

Пример

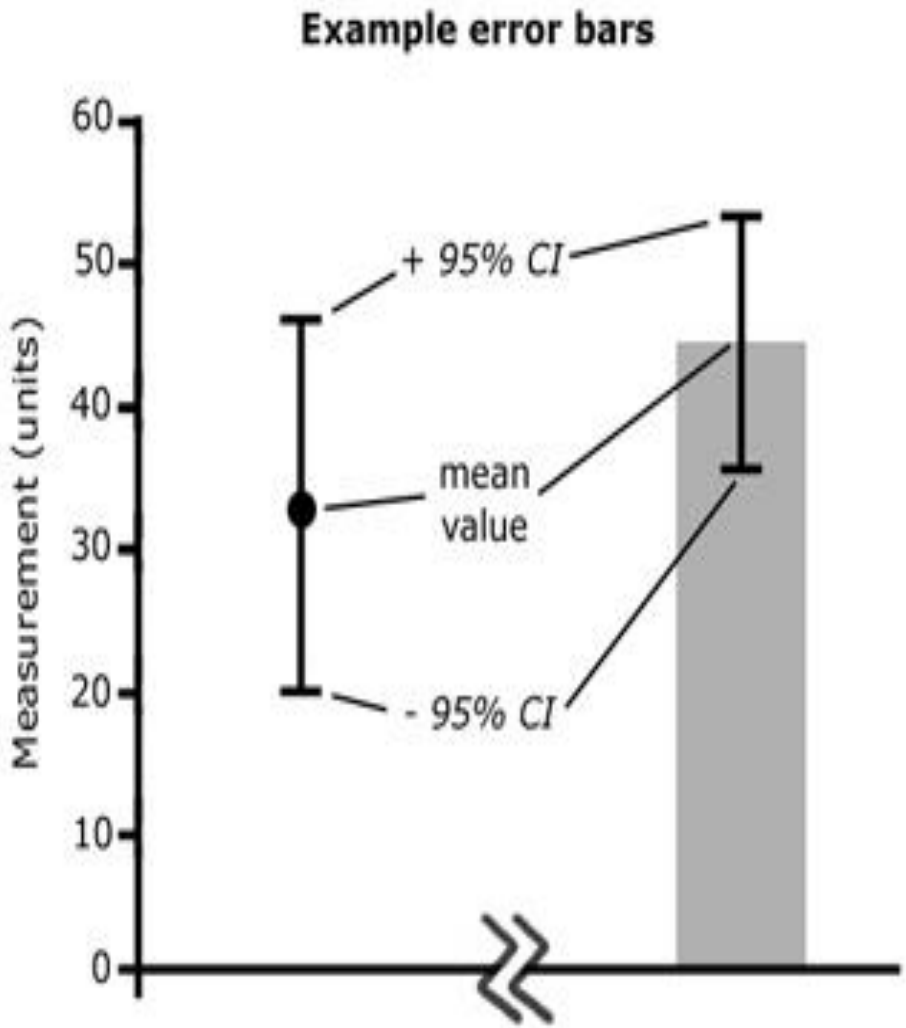
Пусть исследуемой величиной является количество ежегодно обратившихся в клинику пациентов в год за последние 5 лет.

В среднем их количество равно 500,

95% -доверительный интервал – (350, 900). Это означает, что с вероятностью 95%, в течение года в клинику обратятся не менее 350 и не более 900 человек.

Используемое сокращение:

ДИ 95 % (CI 95%) – это доверительный интервал с уровнем доверия 95%.


Example error bars
Measurement (units)
60
50
40
30
20
10
0
+ 95% CI
mean
value
- 95% CI

- В описаниях результатов медико-биологических экспериментов часто используют одно из двух представлений результатов.
- Первое – в виде **« *x* ± SD »**, *где*

  x – среднее,

  SD – *стандартное* отклонение .
- Второе представление результатов – в виде **« *x* ± *m* »**, *где*

  x – среднее

  m – стандартная ошибка среднего (Standard Error of Mean)
- В каждом конкретном случае, необходимо оговаривать, какое из представлений результатов используется, так как запись «одно число плюс/минус другое» может толковаться неоднозначно.

V этап: Анализ результатов исследования

- На этом этапе важное значение имеет качественный логический анализ полученных результатов.
- Использование формальных методов обработки данных без должного логического анализа собранного материала и полученных результатов может привести к неправильным выводам.

Примеры клинических исследований

- Изучение эффективности нового лекарства
- Оценка нового диагностического теста
- Сравнительный анализ схем ведения больного
- Изучение причин и факторов риска болезни
- Прогноз развития заболевания

Прежде чем приступить к анализу данных и проверке различных гипотез:

1. Сформулируйте вопрос, на который Вы хотите ответить с помощью статистического анализа.

2. Выберите наиболее адекватный для ответа на данный вопрос статистический критерий или метод.

3. Правильно интерпретируйте его результаты.

Анализ организации конкретного исследования и его результатов

- Оценить адекватность дизайна научного исследования решению той или иной проблемы эпидемиологии и общественного здоровья.
- Анализ технологии приведенного исследования.
- Оценка полученных результатов.
- Практическое применение полученных результатов.

Некоторые направления статистического анализа

- Сравнение двух и более групп результатов и определение значимости различия
- Выявление и измерение взаимосвязи между признаками
- Изучение динамики явления
- Анализ выживаемости
- Анализ прогностических факторов

Сравнение и определение достоверности различия

независимые группы

опыт-контроль

сопряженные группы

«до» и «после»

две группы

более двух групп

параметрические

непараметрические

- Общие принципы сравнения совокупностей основываются на анализе так называемой **нулевой гипотезы (Но).**
- Согласно этой гипотезе первоначально принимается, что между совокупностями (показателями) различия случайны (не достоверны) , т.е., что обе группы вместе составляют один и тот же однородный материал, одну совокупность.
- Статистический анализ должен привести или к отклонению Но-гипотезы, если доказана достоверность полученных различий, или к ее сохранению, если достоверность различий не доказана, т.е. различия признаны случайными.

- Т.к. статистические различия всегда характеризуются определенным уровнем значимости, то принятие решения по отбрасыванию или сохранению Но - гипотезы связано с оценкой уровня значимости.
- В медико-биологических исследованиях общепризнанным минимальным уровнем значимости является p=0.05.
- Если при сравнении совокупностей полученный при исследовании уровень значимости **меньше 0,05 (p < 0,05**), то Но-гипотеза отбрасывается и **различия в совокупностях признаются достоверными,** воспроизводящимися при повторных исследованиях с определенной вероятностью;
- Если при сравнении совокупностей полученный в исследовании уровень значимости **больше 0,05 (p > 0,05),** то Но- гипотеза признается верной (т.е.подтвержденной), что свидетельствует об **отсутствии достоверных различий** между совокупностями.

Тесты для сравнение двух независимых групп

- Нормальное распределение (параметрические тесты):

  t-тест (Стьюдента)

- Иное распределение

  (непараметрические тесты):

  тесты Колмогорова-Смирнова, Манна-Уитни, Ван-дер-Вардена, Вальда-Вольфовица

Параметрические тесты

- Параметрические методы оценки требуют знания характера распределения (только для нормального распределения) изучаемого признака и его параметров (средних величин, стандартного отклонения и др.).
- Уровень значимости в этих методах определяется с помощью расчета критерия t и сравнения его значения с табличным, который соответствует определенному уровню значимости:

- если $t_{ф} > t_{0,05}$, то $p < 0,05$ и Но - гипотеза отвергается;
- если $t_{ф} < t_{0,05}$, то $p > 0,05$ и Но- гипотеза принимается,

$t_{ф}$ - фактический критерий, рассчитанный в исследовании;

$t_{0,05}$ - табличное значение критерия t для $p = 0.05$.

t – тест (Стьюдента) при сопоставлении двух независимых групп (опыт-контроль):

для частотных показателей:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^{\ 2} + m_2^{\ 2}}}$$

для средних величин:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{m_1^{\ 2} + m_2^{\ 2}}}$$

- Для оценки достоверности полученного критерия «t» при числе наблюдений больше 30 можно пользоваться следующей закономерностью:
- если критерий $t \geq 2$, то он достоверен, т.к. соответствует $p \geq 0{,}95$ или $p \leq 0{,}05$
- если критерий $t \geq 3$, то он достоверен с большей степенью достоверности, т.к. соответствует $p \geq 0{,}99$ или $p \leq 0{,}01$.

- При оценке двух методов операции в двух группах больных (n_1= 145; n_2=147) в качестве критерия была взята средняя длительность послеоперационного периода. Необходимо оценить достоверность различия по этому критерию. (Предполагается нормальное распределение изучаемого признака.)
- Средняя длительность послеоперационного периода в соответствующих группах больных:

  для метода №1: 9 дней, m_1=0,3 дн.

  для метода №2: 11 дней, m_2=0,2 дн.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{m^2{}_1 + m^2{}_2}} = \frac{11 - 9}{\sqrt{0{,}3^2 + 0{,}2^2}} = 5{,}0$$

- Так как n > 30 для оценки достоверности критерия t можно использовать следующую закономерность: $t_{0,05}$ = 2; $t_{0,01}$ = 3.
- **<u>Вывод:</u>** Т.к. $t_{факт.}(5{,}0) > t_{0,01}(3)$, следовательно различия в средней длительности послеоперационного периода достоверны ($p < 0{,}01$), и по этому показателю метод №1 **статистически значимо** лучше метода №2.

- Для оценки эффективности вакцинации против гриппа провели изучение заболеваемости среди привитых и непривитых. Необходимо оценить достоверность различия между этими показателями.
- Заболеваемость непривитых: P_1=13,2 ‰, m_1=0,9 ‰
- Заболеваемость привитых: P_2=10,6 ‰, m_2=1,1 ‰

$$t = \frac{P_1 - P_2}{\sqrt{m^2{}_1 + m^2{}_2}} = \frac{13{,}2 - 10{,}6}{\sqrt{1{,}1^2 + 0{,}9^2}} = 1{,}8$$

Так как $n_{1,2}$ > 30 для оценки достоверности критерия t можно использовать следующую закономерность:

$t_{0,05}$ = 2; $t_{0,01}$ = 3.

<u>Вывод</u>: Т.к. $t_{факт.}$(1,8) < $t_{0,05}$ (2), следовательно различие в уровнях заболеваемости гриппом среди привитых и непривитых **статистически незначимо**, и нет оснований считать противогриппозную вакцину эффективной.

Тесты для сравнения двух сопряженных групп («до и после»)

- Нормальное распределение (параметрический тест)

  парный t-тест

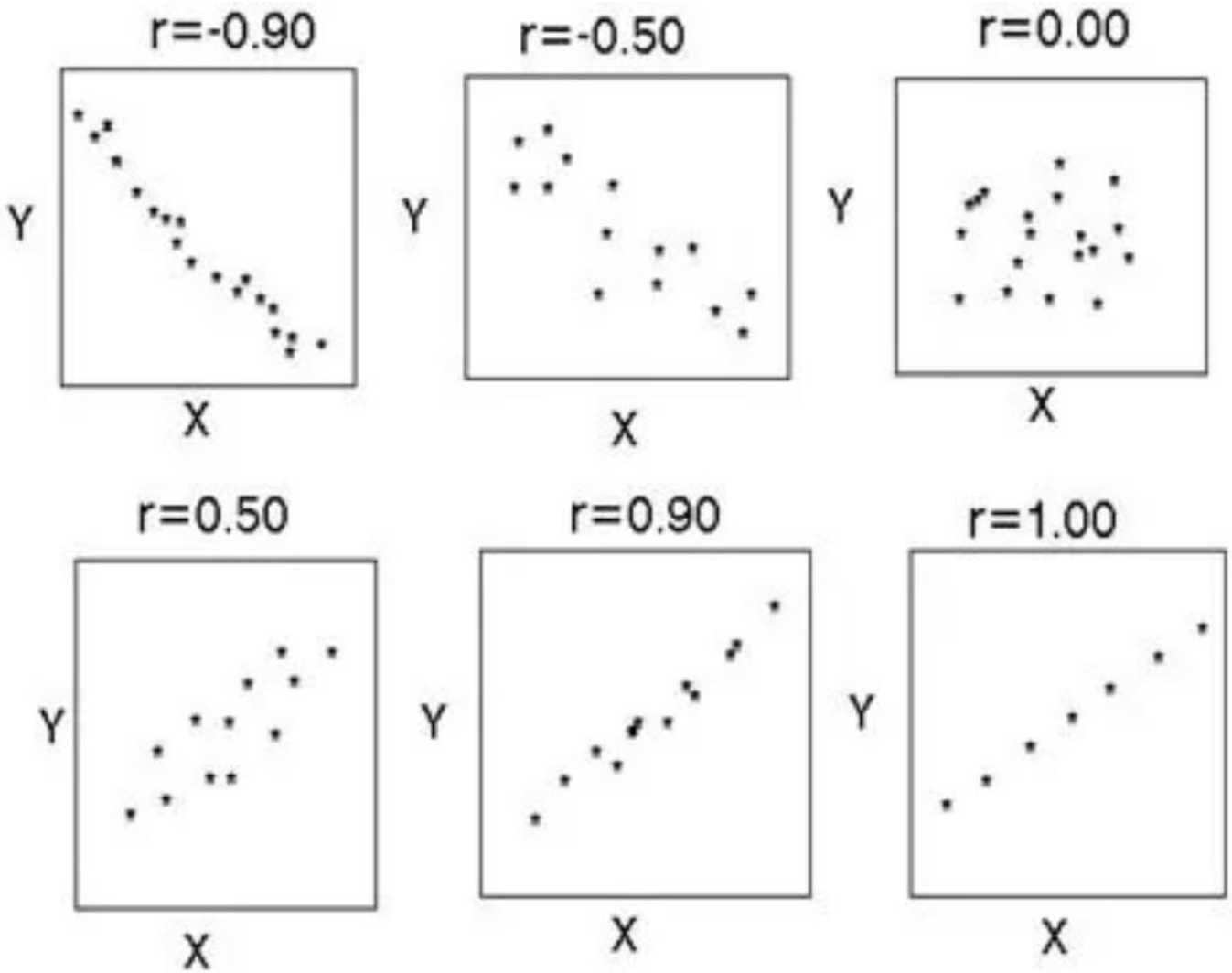
- Отличное от нормального распределение (непараметрические тесты)

  тесты знаков, Вилкоксона

МЕТОДЫ АНАЛИЗА ВЗАИМОСВЯЗИ МЕЖДУ КАЧЕСТВЕННЫМИ И КОЛИЧЕСТВЕННЫМИ ПРИЗНАКАМИ

Корреляционный анализ устанавливает:

- **наличие связи;**
- **силу связи**: слабая (коэффициент корреляции до 0.29), средняя (0.3 - 0.69), сильная (0.7 и выше);
- **направление связи**: прямая (изменения признаков происходят в одном на-правлении) и обратная (изменения признаков происходят в разных направлениях);
- **характер связи**: парциальная (частная) (взаимосвязь между парой признаков) и множественная (взаимосвязь группы признаков).


r=-0.90
Y
X
r=-0.50
Y
X
r=0.00
Y
X
r=0.50
Y
X
r=0.90
Y
X
r=1.00
Y
X

Примеры

- Между САД и ДАД у 18-летних юношей существует статистически значимая прямая сильная корреляционная связь (r_{xy} = 0,7, p = 0,032).
- Между частотой заболеваний нервной системы у работающих в условиях с различной интенсивностью шума и интенсивностью шума существует статистически значимая прямая сильная корреляционная связь (r_{xy} = 0,8, p = 0,043).
- Между частотой язвенной болезни желудка у хирургов и стажем работы по специальности не существует статистически значимой корреляционной связи (r_{xy} = 0,3, p = 0,084).

- **Регрессионный анализ** - метод статистической обработки данных, позволяющий по средней величине одного признака определить среднюю величину другого признака, корреляционно связанного с первым.

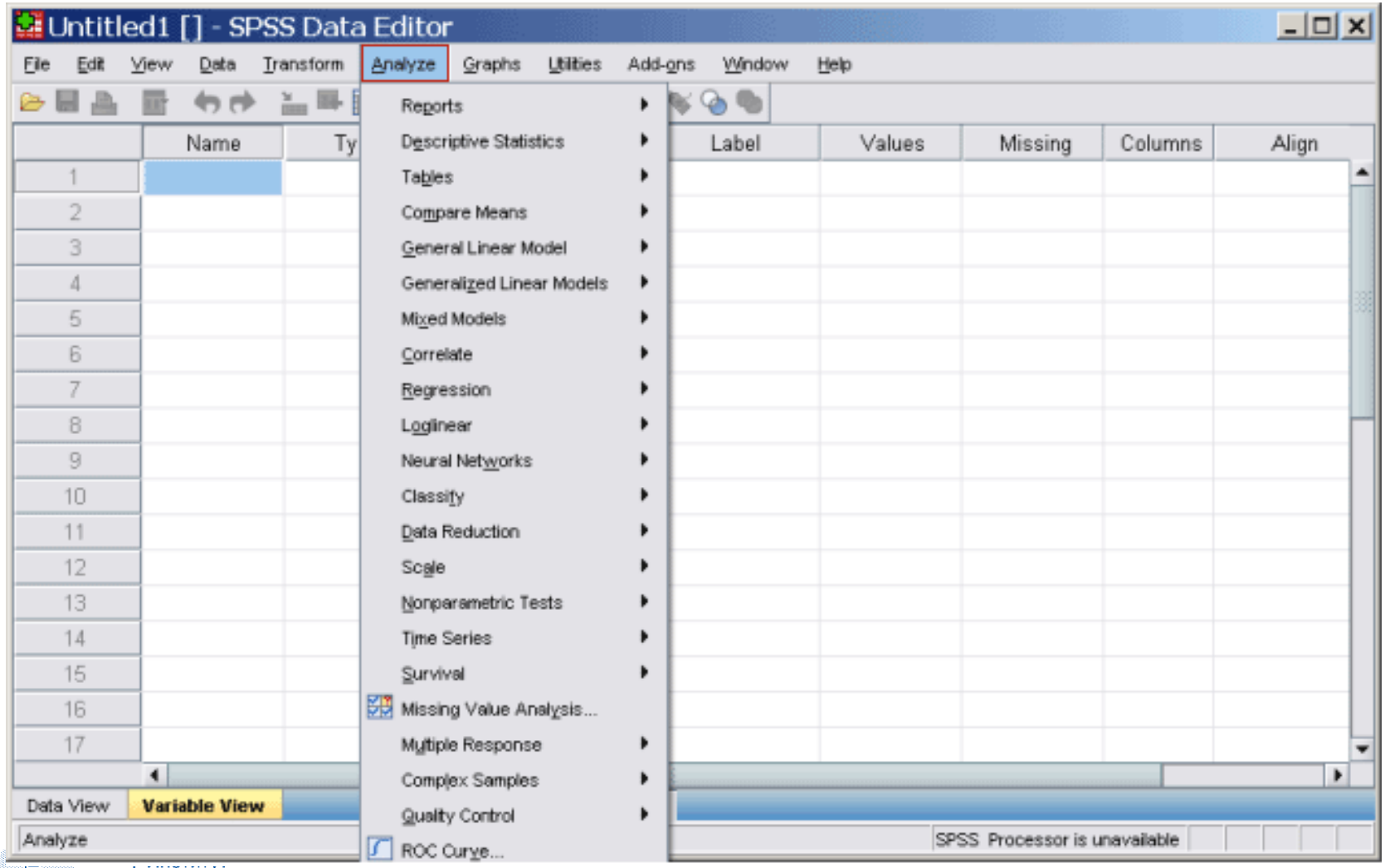
- **Знание возможностей статистических методов необходимо каждому работающему в медицине и здравоохранении.**

	Type of Data			
Goal	**Measurement (from Gaussian Population)**	**Rank, Score, or Measurement (from Non- Gaussian Population)**	**Binomial (Two Possible Outcomes)**	**Survival Time**
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample *t*-test	Wilcoxon test	Chi-square or Binomial test **	
Compare two unpaired groups	Unpaired *t*- test	Mann-Whitney test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel*
Compare two paired groups	Paired *t*- test	Wilcoxon test	McNemar's test	Conditional proportional hazards regression*
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression**
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**	Conditional proportional hazards regression**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**	
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*	Cox proportional hazard regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**		Multiple logistic regression*	Cox proportional hazard regression*

Пакеты прикладных статистических программ

- SPSS (SPSS Inc.,USA)
- SAS
- STATA
- STATISTICA (StatSoft, USA; StatSoft-Russia)
- BIOSTATISTICA (S.A. Glantz, McGraw Hill, перевод на русский язык – «Практика», 1998)
- EpiInfo 2000 (Centers for Disease Control and Prevention, USA)
- R

SPSS - Statistical Package for Social Science - Статистический пакет для социальных наук



ЗАДАНИЯ ДЛЯ АСПИРАНТОВ В ЧЕК-ЛИСТ ПО МЕТОДАМ СТАТИСТИЧЕСКОГО АНАЛИЗА (лекция 2) ПО МОДУЛЮ «МЕТОДОЛОГИЯ НАУЧНОГО ИССЛЕДОВАНИЯ»

Какое из приведенных ниже требований к выборочной совокупности является основным:

А. Однородность
Б. Типичность
В. Репрезентативность
Г. Достаточность
Д. Комплексность

Под доверительным интервалом понимают:

А. Интервал, в пределах которого находится показатель в генеральной совокупности
Б. Интервал, в пределах которого колеблется средняя арифметическая в вариационном ряду
В. Интервал, которому доверяют авторы исследования
Г. Интервал, в пределах которого мониторируется показатель
Д. Интервал, в пределах которого могут колебаться статистические критерии

Основным условием применения параметрических методов анализа является:

А. Формирование случайной выборки
Б. Наличие двух независимых выборок
В. Корреляционная связь между признаками
Г. Невозможность применения непараметрических методов
Д. Нормальное распределение признака

К непараметрическим критериям в статистике относят все нижеперечисленные, кроме:

А. Критерий Стьюдента

Б. Критерий Вилкоксона

В. Критерий Манна-Уитни

Г. Критерий Колмогорова-Смирнова

Как вы понимаете статистическую значимость полученных результатов исследования

А. Как воспроизводимость результатов, полученных в ходе выборочного исследования, в генеральной совокупности с вероятностью $p \geq 0,95$

Б. Как отсутствие систематических ошибок и ошибок регистрации при проведении исследования

В. Как абсолютную точность результатов, полученных на основании применения надежных и современных методик исследования.

Сделайте вывод: значимо ли влияние лечения, проведенного больным, на уровень артериального давления, если до лечения АД: 195±3.0/121±3.0 мм рт.ст, а после - АД: 140±2.7/90±2.7 мм рт.ст, (p=0,065)

А) влияние статистически значимо

Б) различие случайно

Коэффициент корреляции массы тела и роста 15-летних девочек r=+0,31. Выберите правильный вывод:

А) существует прямая, слабая взаимосвязь между массой тела и ростом девочек

Б) при увеличении среднего роста девочек на 1см средняя масса их увеличивается на 0,31 кг

В) для определения роста девочек надо массу умножить на 0,31

Оцените возможность рекомендации вакцины против гриппа для широкого применения, если среди привитых заболеваемость гриппом составила 20,0%±3,0%, а среди непривитых –35,0%±7,0% (p=0,035)

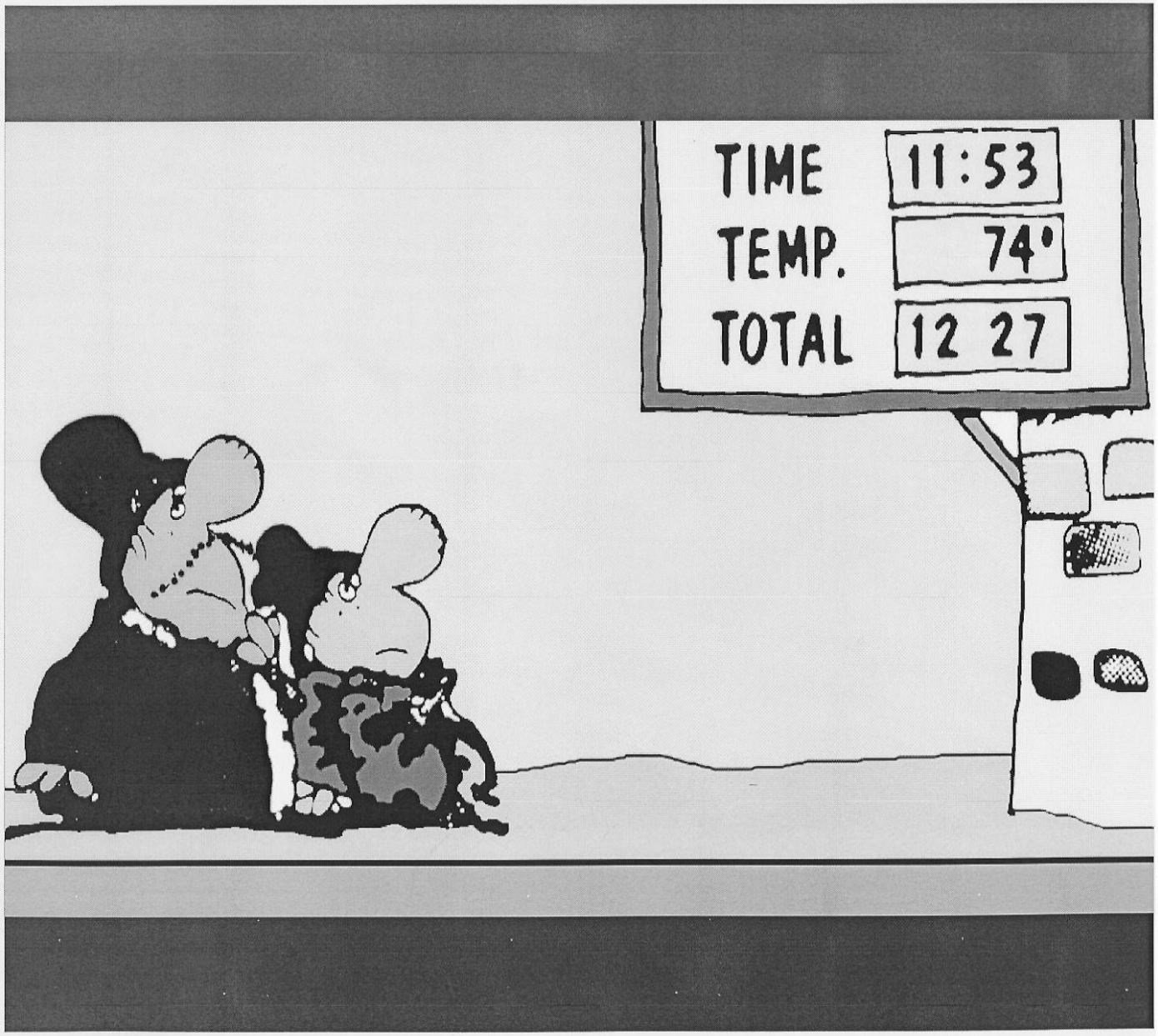
А) эффективна, рекомендовать

Б) не эффективна, не рекомендовать

В) при данных условиях оценку эффективности провести нельзя

Приведены результаты выборочного исследования: при анализе физического развития 225 мальчиков, обучающихся в одной из средних школ г.Н, у 8,0% из них было ожирение.

Проведите точечную и интервальную оценку параметра выборочной совокупности. Сделайте соответствующие выводы.


TIME
11:53
TEMP.
74°
TOTAL
12 27

Спасибо!
Вопросы?