

Описательная статистика

И.И.Косаговская

- Неверно организованный эксперимент не спасет никакой статистический анализ.

**Что представляет из себя исследователь,
закончивший сбор материала?**



**Какой представляется исследователю
обработка результатов?**



Основные этапы статистического анализа

- Описание полученного массива данных**
- Анализ данных и проверка различных гипотез**

Описание полученного массива данных

Descriptive Statistics

Прежде чем приступать к описанию признака, определите его тип.

ВНИМАНИЕ !



**От типа признака зависит выбор
статистического пути его описания
(обобщения)**

**Признаки, или переменные (variables),
могут принимать различные
конкретные значения (values).**

Типы признаков (виды шкал)

- Качественные, категориальные (qualitative, categorical)
 - Номинальные (Nominal) (частный случай : бинарные, дихотомические (Binary – dichotomous))
 - Порядковые, ординальные, ранжируемые (Ordinal)
- Количественные, интервальные (quantitative, numerical, interval)
 - Дискретные (Discrete)
 - Непрерывные (Continuous)

Описательная статистика

- занимается представлением и описанием данных и включает:
 - Методы представления данных (таблицы, гистограммы и т.д.)
 - Описание массива данных

Описание массива данных

Номинальные и порядковые (ординальные) признаки описываются (обобщаются) путем расчета доли (пропорции, относительной частоты)

Описание массива данных

- Единственный способ описать качественные признаки заключается в расчете доли от общего числа объектов (или пропорции), которая приходится на то или иное значение.

Доля может быть выражена в процентах:

$$P = \frac{n_1}{n_1 + n_2} \times 100\% = \frac{n_1}{n} \times 100\%$$

где n_1 и n_2 – численности групп (имеющих и не имеющих изучаемый признак), а $n = n_1 + n_2$ – численность всей совокупности.

Описание массива данных

- Эти величины чаще всего используются для характеристики структуры изучаемой совокупности или оценки частоты изучаемого явления в популяции.
- Масштабирующим коэффициентом может быть 100 (%), 1000 (‰), 10 000 (‱), 100 000 (‵) (‰, ‱, ‵)

Пример

- Был выделен 21 кишечный паразит при обследовании детей:

Giardia lamblia

Entamoeba histolytica

Ascaris lumbricoides

Enterobius vermicularis

Ascaris lumbricoides

Enterobius vermicularis

Giardia lamblia

Giardia lamblia

Entamoeba histolytica

Ascaris lumbricoides

Enterobius vermicularis

Ascaris lumbricoides

Enterobius vermicularis

Giardia lamblia

Giardia lamblia

Entamoeba histolytica

Ascaris lumbricoides

Enterobius vermicularis

Ascaris lumbricoides

Enterobius vermicularis

Giardia lamblia

Пример

Визуальное упорядочивание

Giardia lamblia

Giardia lamblia

Giardia lamblia

Giardia lamblia

Giardia lamblia

Giardia lamblia

Ascaris lumbricoides

Ascaris lumbricoides

Ascaris lumbricoides

Ascaris lumbricoides

Ascaris lumbricoides

Ascaris lumbricoides

Enterobius vermicularis

Enterobius vermicularis

Enterobius vermicularis

Enterobius vermicularis

Enterobius vermicularis

Enterobius vermicularis

Entamoeba histolytica

Entamoeba histolytica

Entamoeba histolytica

Пример

- Частотное распределение

Parasite	n
<i>Giardia lamblia</i>	6
<i>Ascaris lumbricoides</i>	6
<i>Enterobius vermicularis</i>	6
<i>Entamoeba histolytica</i>	3
Total	21

Пример

- Распределение относительных частот (долей, пропорций)

Parásito	n	%
<i>Giardia lamblia</i>	6	28.57
<i>Ascaris lumbricoides</i>	6	28.57
<i>Enterobius vermicularis</i>	6	28.57
<i>Entamoeba histolytica</i>	3	14.29
Total	21	100.00

Описание массива данных

Описание (обобщение) количественного признака:

1. Оценка центральной тенденции
2. Оценка разнообразия (разброса, рассеяния).

ВНИМАНИЕ



От вида распределения зависит выбор статистического пути описания (обобщения) и анализа количественного признака

Вид распределения

Под **видом распределения** случайной величины понимают соответствие, устанавливаемое между всеми возможными числовыми значениями случайной величины и вероятностями их появления в совокупности.

Вид распределения

Вид (закон) распределения может быть представлен:

- аналитической зависимостью в виде формулы;
- в виде графического изображения;
- в виде таблицы

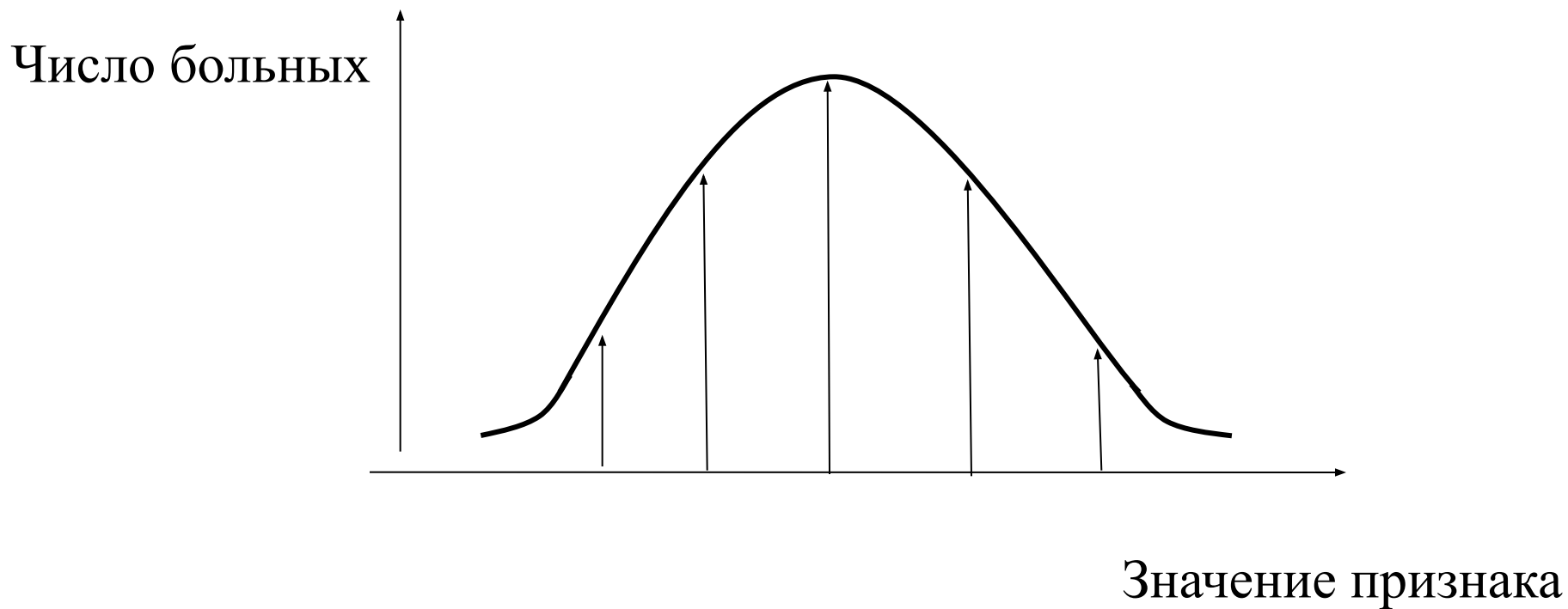
Виды распределения

Нормальное (гауссово, симметричное, колоколообразное) распределение (**normal, Gaussian distribution**)– описывает совместное воздействие на изучаемое явление небольшого числа случайно сочетающихся факторов (по сравнению с общей суммой факторов), число которых неограничено велико.

Встречается в природе наиболее часто, за что и получило название «нормального».

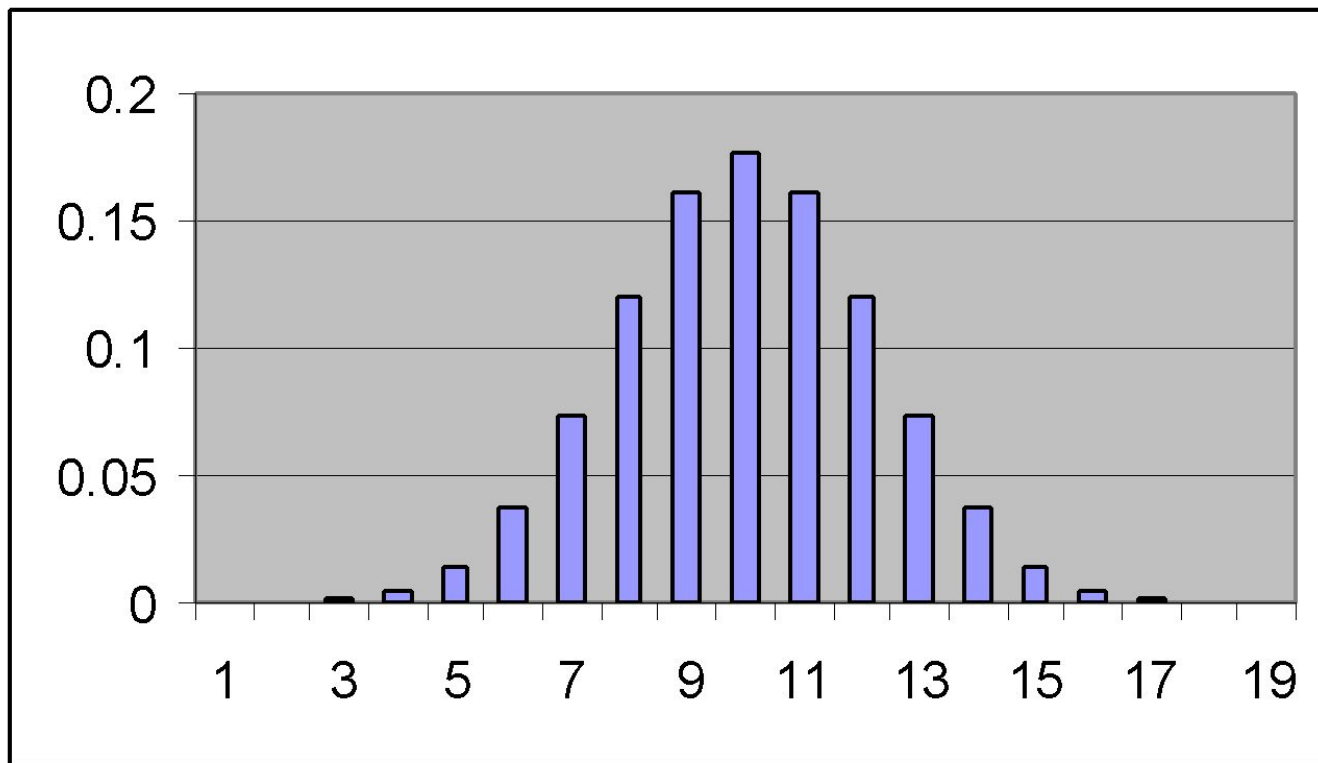
Характеризует распределение **непрерывных** случайных величин.

Кривая нормального распределения



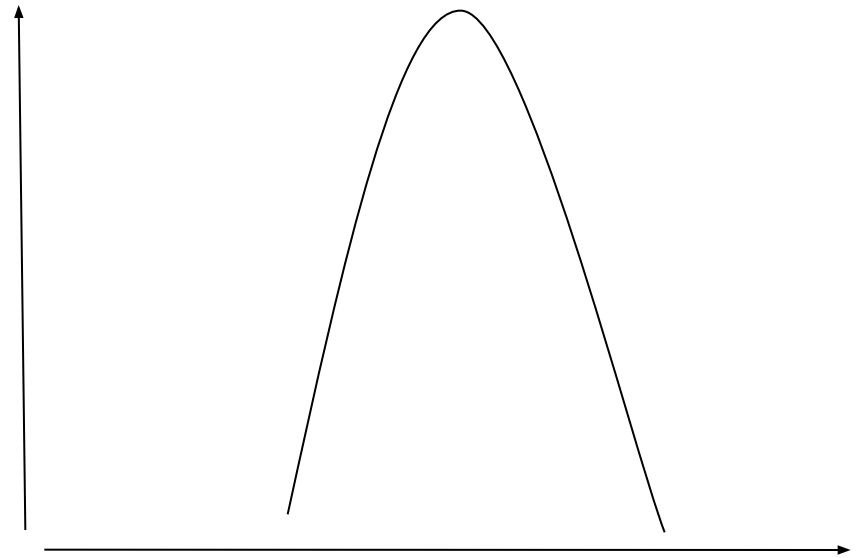
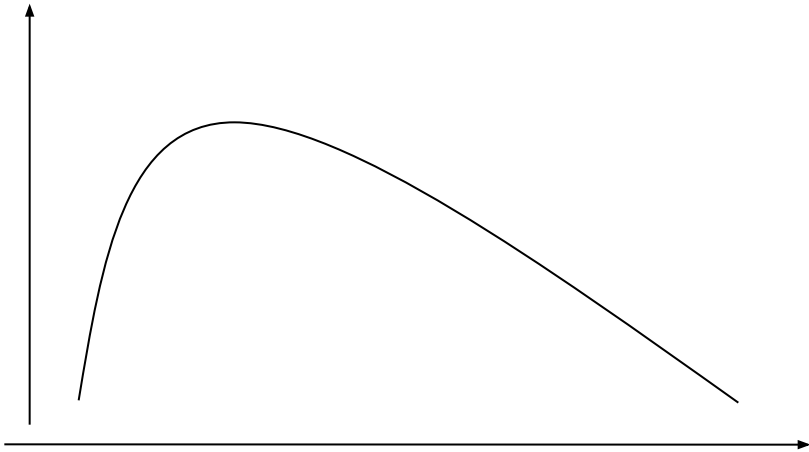
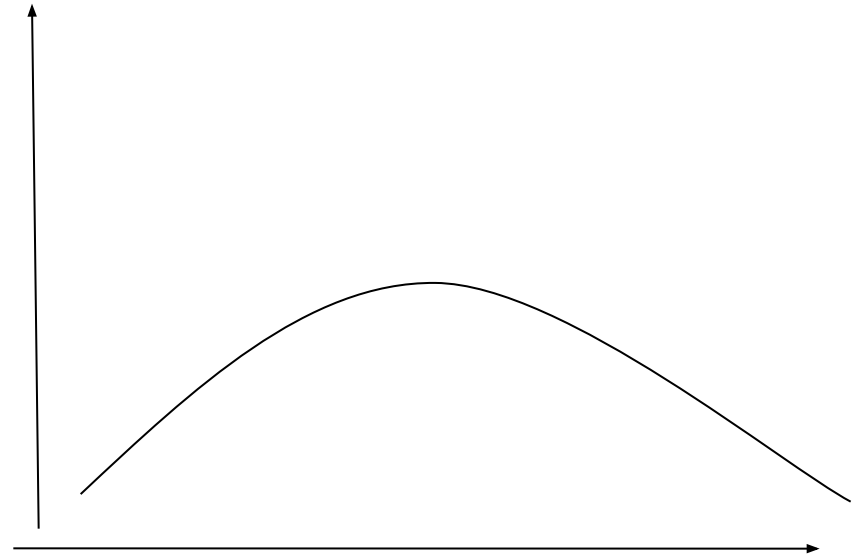
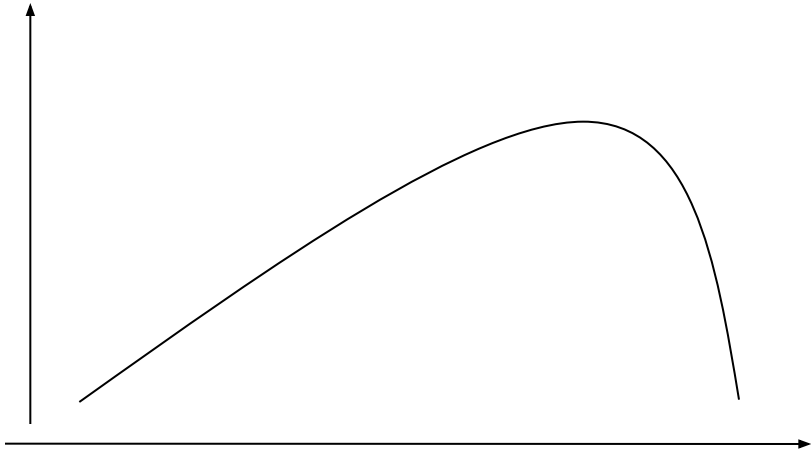
- **Биномиальное (Бернулли) распределение (binomial, Bernoulli distribution)** – описывает распределение частоты события, обладающего постоянной вероятностью появления при многократных испытаниях.
- При большом числе испытаний стремиться к нормальному.

Биномиальное распределение

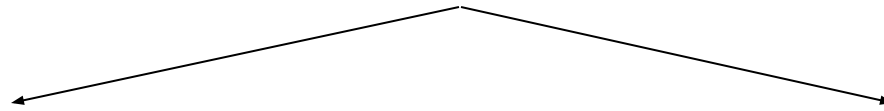


- Крайним вариантом биномиального распределения является **альтернативное** распределение, при котором вся совокупность распределяется на две части (две альтернативы).
- Биномиальное распределение характеризует распределение **дискретных** случайных величин.

- **Распределение Пуассона** – описывает события, при которых с возрастанием значения случайной величины, вероятность появления ее в совокупности резко уменьшается.
- Распределение Пуассона характерно для редких событий и может рассматриваться также как крайний вариант биномиального. Характеризует распределение **дискретных** случайных величин.



Вид распределения



нормальное

отличное от
нормального



Параметрическая
статистика

Непараметрическая
статистика

Непараметрические методы:

- не требуют предварительного знания вида распределения;
- не требуют предварительного расчета параметров распределения (средних величин, стандартного отклонения и др.);
- позволяют сравнивать совокупности с номинальными и порядковыми признаками;
- просты в применении.

Отрицательные стороны непараметрических методов:

- обладают меньшей мощностью, чем параметрические;
- имеют существенные ограничения в применении по числу наблюдений

Вариационный ряд (frequency table)-

ранжированный ряд распределения по величине какого-либо признака.

Этот признак носит название варьирующего, а его отдельные числовые значения называются вариантами и обозначаются через "х".

Число, показывающее, сколько раз данная варианта встречается в вариационном ряду, называется частотой и обозначается через "р"

Результаты измерения частоты пульса у некурящих студентов-медиков в возрасте 20 лет:

68,58,65,55,70,62,60,65,70,58,62,58,62,60,60,65,62,55,
62,58,60,70,62,65,60,68,65,62,68,65,60,62,60,68,65,60,
62,60,65,62,68

Построим вариационный ряд:

x	55	58	60	62	65	68	70
p	2	4	9	10	8	5	3

Вариационный ряд можно разбивать на отдельные части, которые называются **квантилями (quantile)**.

Название квантилей	Число частей, на которые разбивается ряд
Медиана	2
Терциль	3
Квартиль	4
Дециль	10
Процентиль	100

Виды вариационных рядов:

В зависимости от вида случайной величины :

- дискретный
- непрерывный

В зависимости от группировки вариантов:

- несгруппированный
- сгруппированный (интервальный)

В зависимости от частоты, с которой каждая варианта встречается в вариационном ряду:

- простой ($p = 1$);
- взвешенный ($p > 1$).

ХАРАКТЕРИСТИКИ (МЕРЫ) ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

Показатели, характеризующие **центральную тенденцию (central tendency)** :

средние величины, медиана, мода

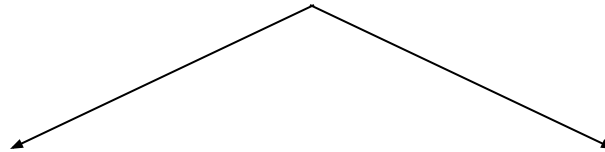
ХАРАКТЕРИСТИКИ РАЗБРОСА (РАЗНООБРАЗИЯ)

Показатели, характеризующие разнообразие (рассеяние, вариацию, разброс) (**spread**) признака:

размах, стандартное отклонение, дисперсия, интерквартильный интервал, коэффициент вариации

- Выбор характеристик центральной тенденции и разнообразия признака прежде всего зависит от вида распределения.

Вид распределения



нормальное

отличное от
нормального



оценка

центральной

тенденции

средняя
арифметическая,
мода, медиана

мода, медиана

- **Средняя величина** - обобщающий коэффициент, который характеризует наиболее типичный размер определенного признака в целом для совокупности или для отдельных ее частей.

- Расчет средних величин имеет смысл только для качественно однородной совокупности, в связи с этим в одной совокупности может быть столько средних, на сколько однородных групп она может быть разбита.

Виды средних величин

Средняя арифметическая (mean) - применяется, если варианты возрастают (убывают) в арифметической прогрессии.

$$\bar{x} = \frac{\sum x_i}{n} \qquad \bar{x} = \frac{\sum (x_i \times p)}{n}$$

\bar{x} - средняя арифметическая;

x_i - варианта;

p - частота встречаемости варианты;

n - число наблюдений

Виды средних величин

- **Средняя геометрическая** - вычисляется, если варианты возрастают (убывают) в геометрической прогрессии

$$\bar{x}_g = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$$

На практике используют логарифмированную формулу:

$$\log \bar{x}_g = 1 / n (\log x_1 + \log x_2 + \log x_3 + \dots \log x_n)$$

Структурные средние

- **Мода (M_o) (mode)**- наиболее часто встречающаяся в вариационном ряду варианта

Мода используется:

- при малом числе наблюдений, когда велико влияние состава совокупности на среднюю ;
- для характеристики центральной тенденции при ассиметричных распределениях, когда велико влияние на среднюю крайних вариантов;

Структурные средние

- **Медиана (Me)(median)** - варианта, которая делит вариационный ряд на две равные части.
- При нечетном количестве значений медиана всегда будет совпадать с одним из измеренных значений. При четном количестве медиана будет средним арифметическим двух соседних значений.

Медиана используется:

- при необходимости знать, какая часть вариант лежит выше и ниже срединного значения ;
- для характеристики центральной тенденции при ассиметричных распределениях .

Пример

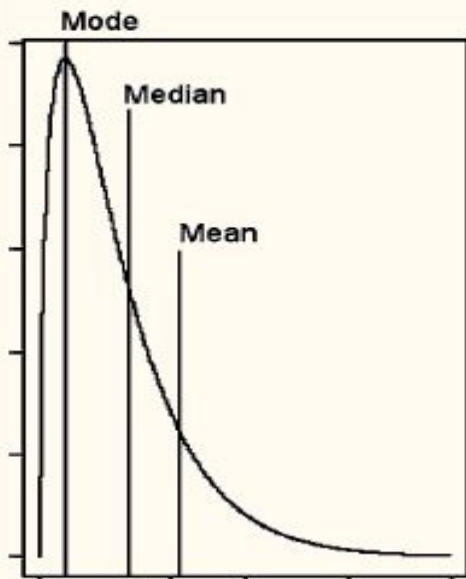
x	55	58	60	62	65	68	70
p	2	4	9	10	8	5	3

$$\bar{x} = \frac{\sum(x_i \times p)}{n}$$

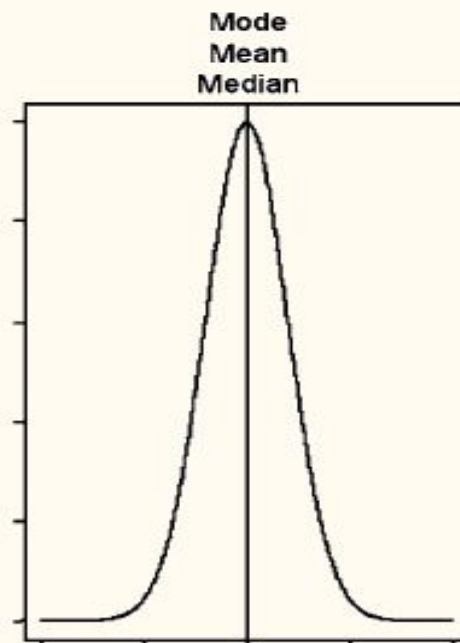
Mean = 62,7 уд.в мин.

Moda = 62 уд.в мин.

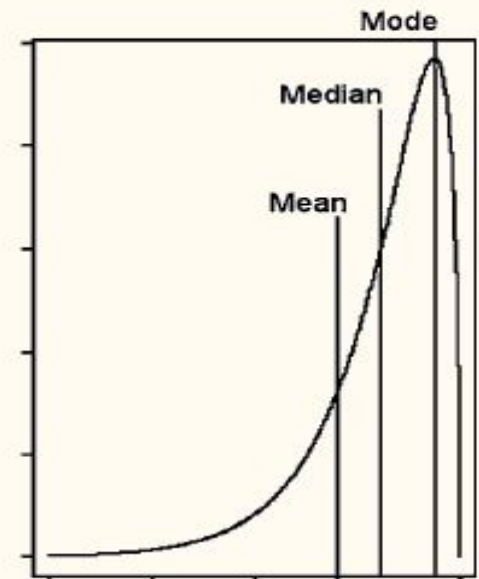
Median = 62 уд.в мин.



Right skew



Normal



Left skew

Средняя может вводить в заблуждение

- Group 1 data: 1,1,1,2,3,3,5,8,20
 - Mean: 4.9 Median: 3 Mode: 1
- Group 2 data: 1,1,1,2,3,3,5,8,10
 - Mean: 3.8 Median: 3 Mode: 1
- Когда объем совокупности небольшой, единичные значения, резко отличающиеся по своей величине от остальных, оказывают большое влияние на размер средней, при этом практически не влияя на моду и медиану.
- В этом случае мода и медиана более информативны, чем средняя.

Характеристики разнообразия вариационного ряда

Размах вариации (амплитуда) (range)

$$A = X_{\max} - X_{\min}$$

$$A = 70 - 55 = 15 \text{ (уд.в мин.)}$$

x	55	58	60	62	65	68	70
p	2	4	9	10	8	5	3

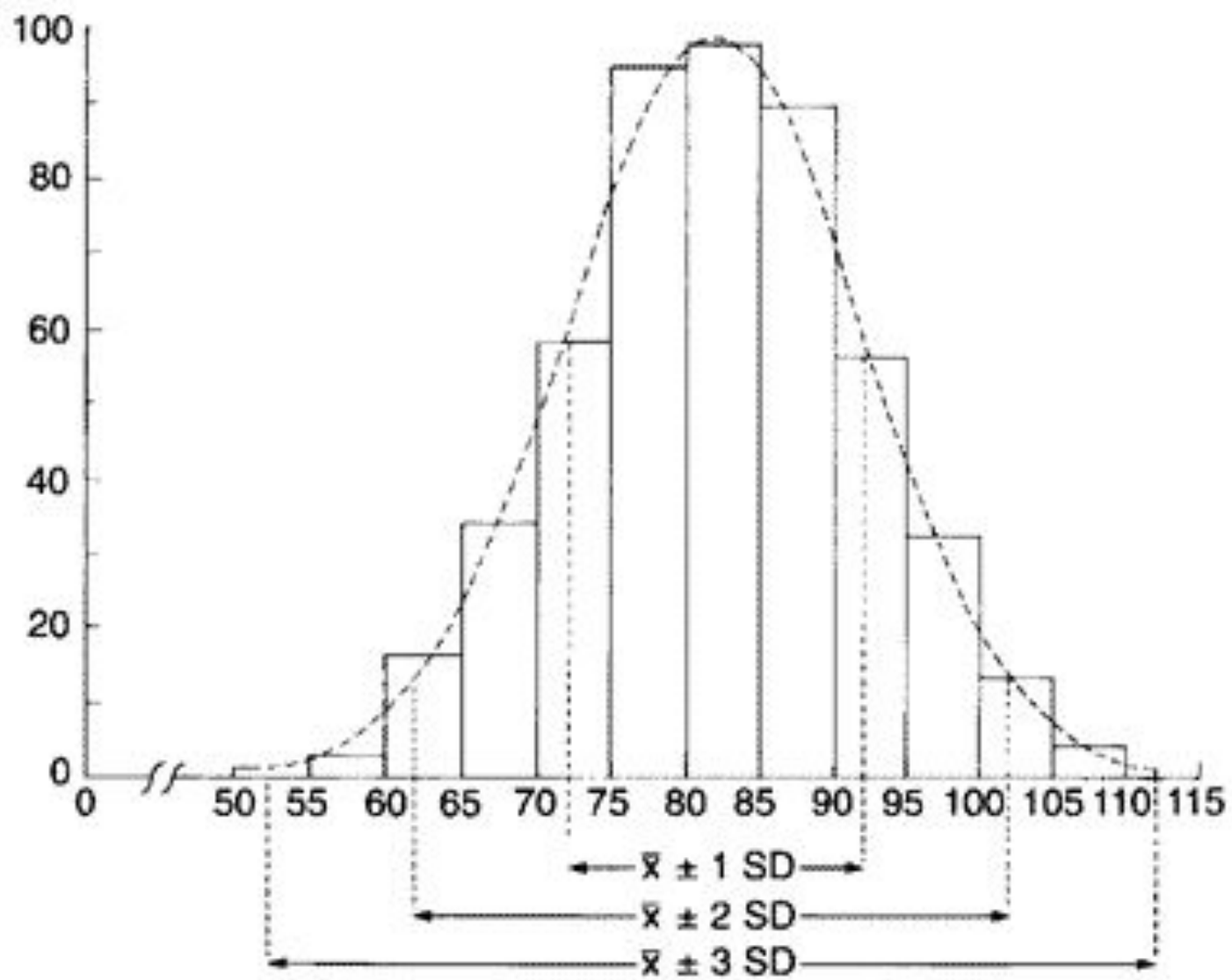
Характеристики разнообразия (разброса)

- Стандартное отклонение (среднее квадратическое отклонение) (standard deviation, SD)

$$\delta \text{ (SD)} = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot p}{n - 1}}$$

Правило трех сигм

- 68.3 % всех вариант отклоняются от средней не более, чем на 1σ ;
- 95.4% вариант находятся в пределах $\bar{x} \pm 2\sigma$;
- 99.7% вариант находятся в пределах $\bar{x} \pm 3\sigma$



Наиболее распространенные ошибки: средняя арифметическая используется для характеристики признаков с «анормальным» распределением или для порядковых признаков

Уровень глюкозы $8,2 \pm 7,5$ ммоль/л

Выраженность боли: $2,5 \pm 1,2$ балла
(1 – слабая, 2 – средняя, 3 – сильная)

- Возраст больных составлял от 18 до 68 лет (средний возраст - $22,8 \pm 4,2$ года).

3 сигмы = 12,6 ; $10,2 \leftrightarrow 35,4$

- Сроки поступления больных составили от 1 до 9 дней (в среднем $2,2 \pm 1,4$ дня).

3 сигмы = 4,2; $-2 \leftrightarrow 6,4$

Характеристики разнообразия (разброса)

- Дисперсия (варианса) (variance)

$$\delta^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot p}{n-1}}$$

Характеристики разнообразия (разброса)

- Коэффициент вариации (variation coefficient)

$$C_v = \frac{\delta}{\bar{x}} \cdot 100 \%$$

Характеристики разнообразия (разброса)

Вариационный ряд

- считается однородным при $C_v < 10\%$,
- обладающим средней вариабельностью (разнообразием) при $C_v = 10-15\%$
- обладающим значительной вариабельностью при $C_v > 15\%$.

Характеристики разнообразия (разброса)

- Коэффициент вариации используется при сравнении вариационных рядов, имеющих различную размерность, или одной размерности, но обладающими резкими различиями в своих значениях, затрудняющими их сопоставление.

Характеристики разнообразия (разброса)

- **Интерквартильный интервал (inter-quartile range, IQR)**

Характеристики разнообразия вариационного ряда

- Вариационный ряд разбивают на четыре интервала, получая, соответственно, 25%, 50% и 75% квантили;
- 25% и 75% квантили называют также нижним (low quartile) и верхним квантилями (high quartile).
- 50% квантиль – это медиана.
- Внутри интерквартильного интервала (между 25% и 75% квантилями) лежат 50% наиболее типичных (близких к центральному) значений.

105 112 125 132 155 158 160 165 168 170 175 185 185 235

$$\text{Median} = \frac{160 + 165}{2} = 162.5$$

105 112 125 132 155 158 160 165 168 170 175 185 185 235

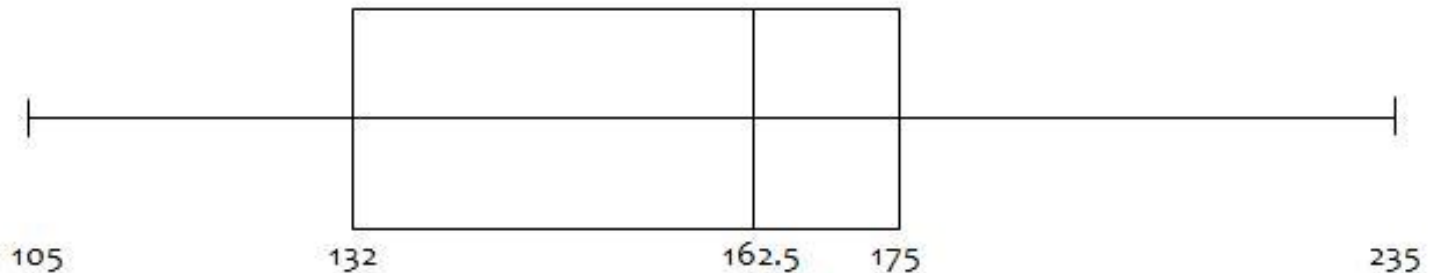
Quartile 1, Q_1 Quartile 3, Q_3

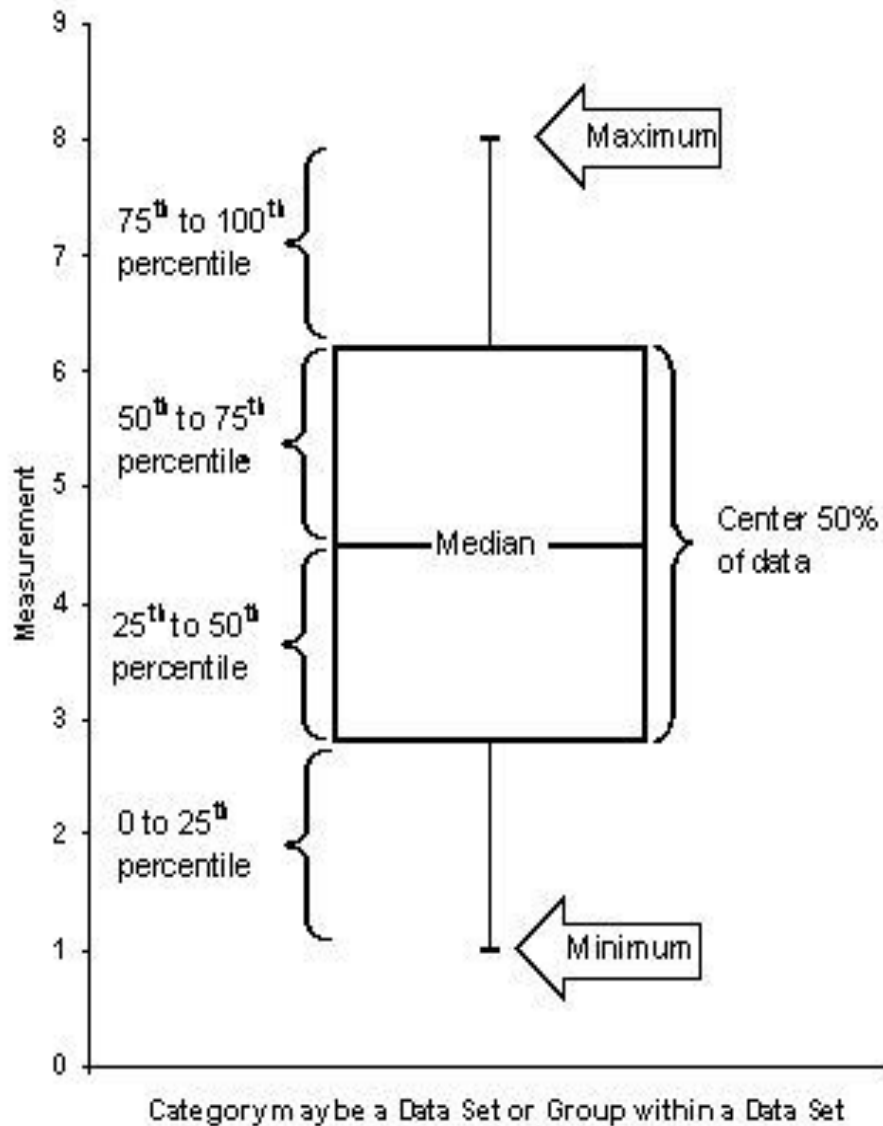
$$\text{Interquartile range (IQR)} = Q_3 - Q_1 = 175 - 132 = 43$$

$$\text{Outlier Test: } 1.5 \times \text{IQR} = 1.5 \times 43 = 64.5$$

$$Q_1 - 1.5 \times \text{IQR} = 132 - 64.5 = 67.5$$

$$Q_3 + 1.5 \times \text{IQR} = 175 + 64.5 = 239.5$$





Первый квартиль (Q1) — это точка на шкале измеренных значений, ниже (левее) которой располагаются 25 % измеренных значений.

Второй квартиль (Q2) — это точка, ниже (левее) которой располагаются 50 % измеренных значений.

Второй квартиль также называется **медианой**.

Третий квартиль (Q3) — это точка на шкале измеренных значений, ниже (левее) которой располагаются 75 % значений.

Пример

- Group 1 data: 1,1,1,2,3,3,5,8,20
 - Mean: 4.9 Median: 3
- Group 2 data: 1,1,1,3,3,3,5,8,10
 - Mean: 3.8 Median: 3
- SDs: group 1: 6.1 group 2: 3.2
- Interquartile range: 1,5

Вид распределения

нормальное

отличное от
нормального

Оценка

разнообразия

(разброса)

стандартное
отклонение
SD

интерквартильный
интервал
IQR

- В случае **нормального распределения** вариационный ряд описывается **средней величиной и стандартным отклонением.**

- Если распределение неизвестно или оно отлично от нормального центральную тенденцию и разброс можно описать с помощью **медианы, нижнего и верхнего квартиля (интерквартильным интервалом)**.

Проверка нормальности распределения

По соотношению средней арифметической, моды и медианы:

- при нормальном распределении, которое обладает симметричностью:

$$\bar{x} = Me = Mo, \text{ или } \bar{x} = \frac{3Me - Mo}{2}$$

- правило "двух третей" Юла: $Mo = 3(Me - 2/3\bar{x})$

Проверка нормальности распределения

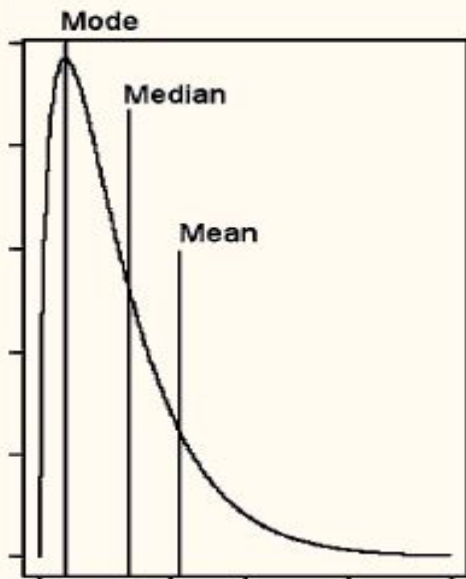
- если распределение симметрично:
 $Me = Mo$
- если распределение обладает правосторонней асимметрией: $Me > Mo$
- если распределение имеет левостороннюю асимметрию: $Me < Mo$

Проверка нормальности распределения

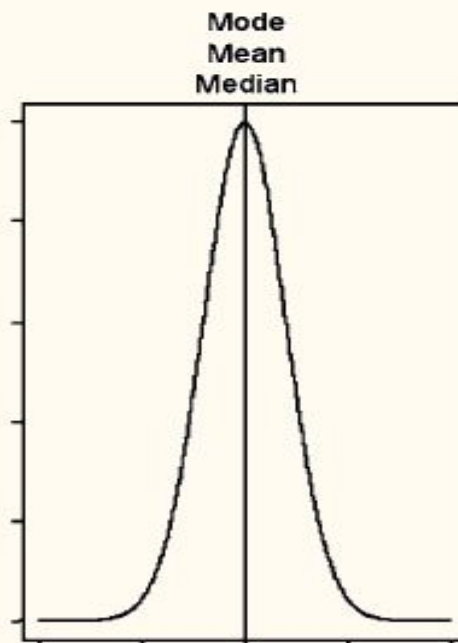
По коэффициенту асимметрии (**skewness**):

$$k_a = \frac{3 \times (\bar{x} - Me)}{\delta}$$

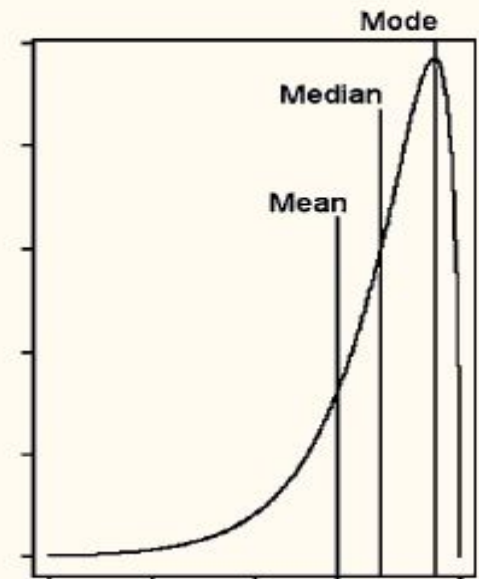
- если распределение симметрично: $k_a = 0$
- при правосторонней асимметрии: $k_a > 0$
- при левосторонней асимметрии: $k_a < 0$



Right skew



Normal



Left skew

Проверка нормальности распределения

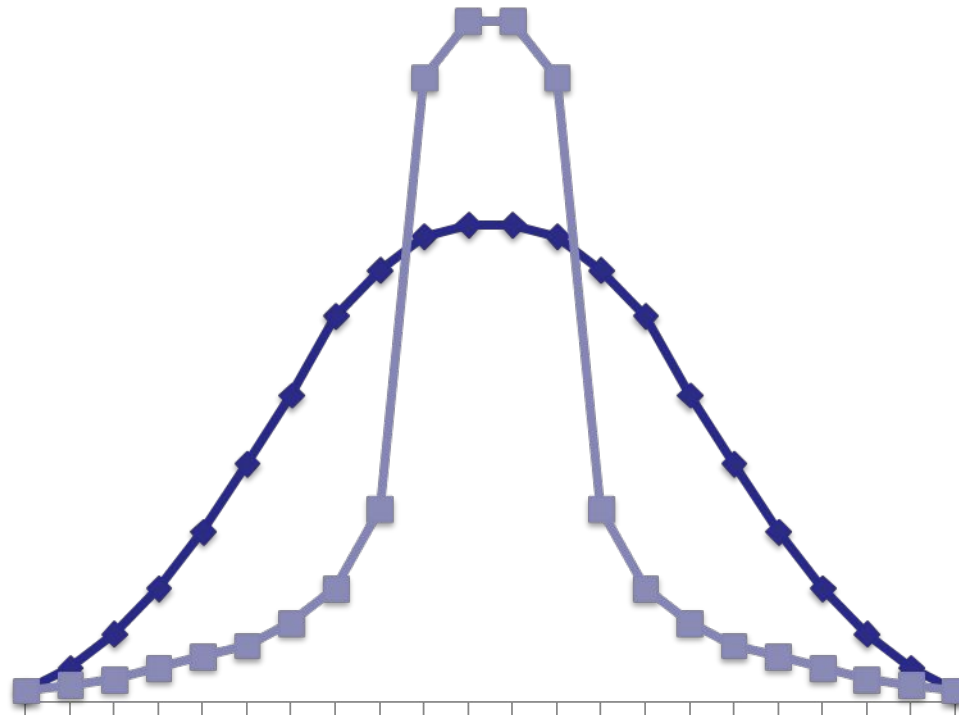
Kurtosis (Коэффициент эксцесса):

- Коэффициент указывает, является ли распределение пологим (при большом значении коэффициента) или островершинным. Коэффициент вариации равен нулю, если наблюдения подчиняются нормальному распределению.
- Если коэффициент вариации значительно отличается от нуля, то гипотезу о том, что данные взяты из нормально распределенной генеральной совокупности, следует отвергнуть.

Вершина более крутая, чем для нормального распределения:
эксцесс положительный, имеются длинные хвосты

распределения;

Вершина положе: **эксцесс отрицательный**, имеются короткие хвосты распределения.



Проверка нормальности распределения

- Если M_e занимает срединное положение между 25-м и 75-м процентилем, то распределение близко к нормальному.

Проверка нормальности распределения

Тесты на нормальность:

- Шапиро-Вилка (Shapiro-Wilk)
- Колмогорова-Смирнова (Kolmogorov-Smirnov)
- Крамера-вон Майса (Kramer-von Mises)
- Андерсона-Дарлинга (Anderson-Darling)

Способы "нормализующего преобразования" (transformation to normality) данных :

- гармоническое преобразование: $1/x$;
- извлечение квадратного корня;
- логарифмирование (дает наиболее точное приближение): $\log x_i$

- Успешность преобразования данных оценивают по коэффициенту асимметрии: чем ближе он к 0, тем ближе экспериментальное распределение к нормальному.

Стандартная ошибка средней

S.E. mean :

- Мера точности выборочной средней (точечная оценка параметра)
- стандартная ошибка позволяет задать доверительный интервал для среднего значения.

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

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.

$$t = \frac{x_{\text{ген}} - x_{\text{выб}}}{m}$$

$$x_{\text{ген}} - x_{\text{выб}} = tm$$

$$\Delta = tm$$

$$x_{\text{ген}} = x_{\text{выб}} \pm \Delta$$

$$x_{\text{ген}} = x_{\text{выб}} \pm tm$$

Доверительный интервал

- Диапазон значений, построенный по выборке, который с определенной степенью доверительности 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 ($\bar{x} \pm 3m$).

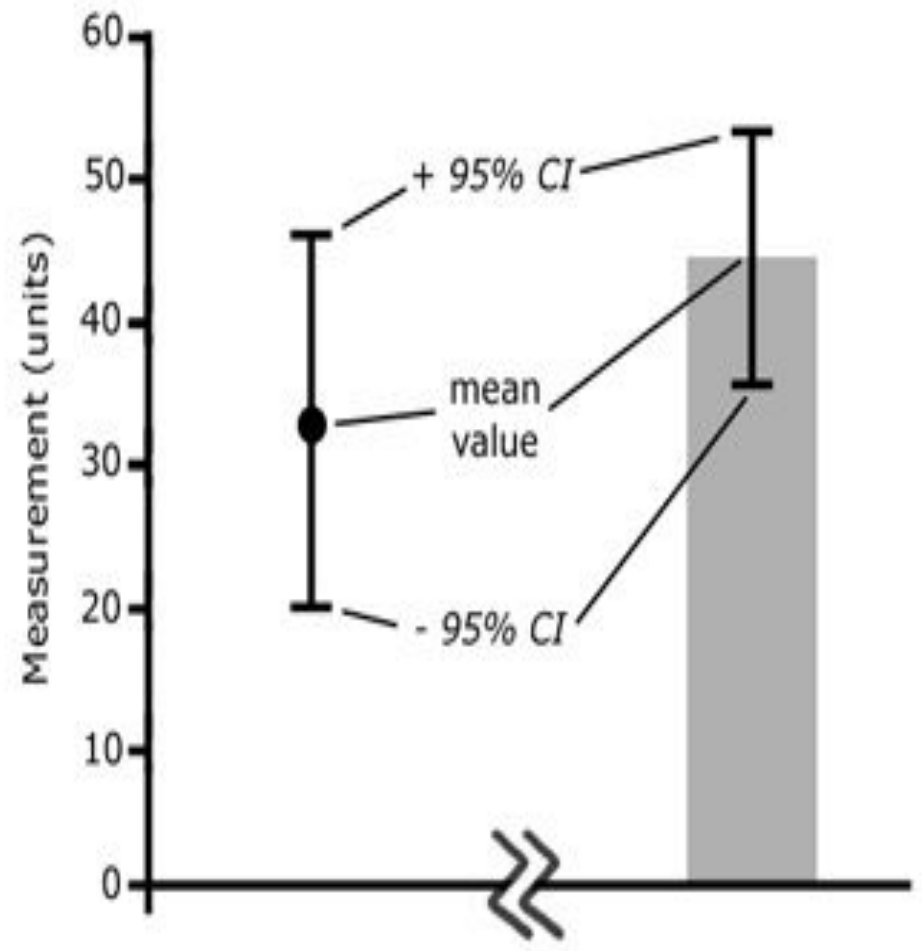
Пример

Пусть исследуемой величиной является количество обратившихся в клинику пациентов в год за последние 5 лет. В среднем их количество равно 500, а 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



- В описаниях результатов медико-биологических экспериментов часто используют одно из двух представлений результатов.
- Первое – в виде « $x \pm SD$ », где x – среднее, а SD – стандартное отклонение .
- Второе представление результатов – в виде « $x \pm m$ », где 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.

Бальная оценка

Порядковая шкала (шкала рангов) – шкала, относительно значений которой уже нельзя говорить ни о том, во сколько раз измеряемая величина больше (меньше) другой, ни на сколько она больше (меньше).

Такая шкала **только упорядочивает объекты**, приписывая им те или иные баллы (результатом измерений является нестрогое упорядочение объектов).

Бальная оценка

Например, так построена **шкала твердости минералов Мооса**: взят набор 10 эталонных минералов для определения относительной твердости методом царапания. За 1 принят тальк, за 2 – гипс, за 3 – кальцит и так далее до 10 – алмаз. Любому минералу соответственно однозначно может быть приписана определенная твердость. Если исследуемый минерал, допустим, царапает кварц (7), но не царапает топаз (8), то соответственно его твердость будет равна 7.

Аналогично построены **шкалы силы ветра Бофорта** и **землетрясений Рихтера**.

Бальная оценка

Шкалы порядка широко используются в педагогике, психологии, медицине и других науках, не столь точных, как, скажем, физика и химия. В частности, повсеместно распространенная **шкала школьных отметок** в баллах (пятибалльная, двенадцатибалльная и т.д.) может быть отнесена к шкале порядка.

Бальная оценка

В медикобиологических исследованиях шкалы порядка встречаются сплошь и рядом и подчас весьма искусно замаскированы.

Например, для анализа свертывания крови используется тромботест: 0 – отсутствию свертывания в течение времени теста (а через минуту?), 1 – «слабые нити», 2 – желеподобный сгусток, 3 – сгусток, легко деформируемый, 4 – плотный, упругий, 5 – плотный, занимающий весь объем и т.п.

Понятно, что интервалы между этими плохо отличимыми и очень субъективными позициями произвольны. В этом случае фраза «Тромботест у исследуемых животных повышался в среднем с 3,3 до 3,7» выглядит абсурдной.

Бальная оценка

Например, можно для оценки степени регенерации суставного хряща после его повреждения применять 100-балльную шкалу. Но она слишком детальна, и ее можно перестроить в пятибалльную («1» – от «1» до «10»; «2» – от «10» до «30» и т.д.), или двухбалльную (например, положительная оценка – все, что выше 70 баллов, отрицательная – 70 и меньше).

Возникает проблема – какие преобразования можно применять к тем или иным типам исходных данных. Другими словами, переход от какой шкалы к какой является корректным. Эта проблема в теории измерений получила название *проблемы адекватности*.

Бальная оценка

Общий вывод – всегда возможен переход от более мощной шкалы к менее мощной, но не наоборот (например, на основании оценок, полученных в шкале отношений, можно строить балльные оценки в порядковой шкале, но не наоборот).

Бальная оценка

В порядковой шкале ничего нельзя сказать о равномерности или неравномерности интервалов между соседними значениями оценок.

Мы не вправе, к примеру, сказать о том, что регенерация хряща, оцененная на «50», настолько же отличается от регенерации, оцененной на «40», как и процессы, оцененные на «90» и «80».

И поэтому совершенно некорректно использование подобных шкал даже в качестве дополнительных аргументов для обоснования эффективности новых методов воздействия на болезнь, поскольку усреднение предполагает сложение значений величины, а операция суммы для порядковых шкал не может быть корректно определена. Соответственно не могут быть определены и все остальные арифметические и алгебраические действия.

Бальная оценка

Поэтому, например, утверждение о том, что регенерация хряща в группе животных с применением нового высокоточного малоинвазивного метода хирургического лечения в среднем на 5,5 баллов выше, чем в группе контрольных животных, будет неправомочным, некорректным.

Тем более при использовании балльных оценок некорректны (даже абсурдны) утверждения типа: «эффективность экспериментальной методики в 1,6 раза выше контрольной».

Бальная оценка

Таким образом, операция вычисления среднего арифметического не является корректной в порядковой шкале.

Шкалу балльных оценок, также как и другие шкалы порядка, можно использовать в экспериментальных исследованиях, но в этом случае необходимо применять адекватные методы обработки данных, не вычисляя «среднего балла».

Корректной характеристикой набора балльных оценок является *медиана* (такое значение оценки, справа и слева от которого расположено одинаковое число оценок в их упорядоченной совокупности).

Бальная оценка

Еще раз повторим – не следует складывать, вычитать, умножать или делить баллы друг на друга, да и на чтобы то ни было – все это абсолютно бессмысленные операции.

В порядковой шкале для усреднения используют медиану

Untitled - SPSS for Windows Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help



11:age		777
	icd7	icd8
1	.	.
2	.	.
3	.	.
4	.	.
5	.	.
6	.	.
7	.	.
8	.	.

- Reports ▶
- Descriptive Statistics ▶**
 - Frequencies...
 - Descriptives...
 - Explore...
 - Crosstabs... cd1c
- Custom Tables ▶
- Compare Means ▶
- General Linear Model ▶
- Correlate ▶
- Regression ▶
- Loglinear ▶
- Classify ▶
- Data Reduction ▶
- Scale ▶
- Nonparametric Tests ▶
- Time Series ▶
- Survival ▶
- Multiple Response ▶
- Missing Value Analysis...

78.59
99.50
79.39
79.35
79.35
79.35
79.35

Summary Report: Statistics



Statistics:

Mean
Median
Grouped Median
Std. Error of Mean
Sum
Minimum
Maximum
Range
First
Last
Standard Deviation
Variance
Kurtosis
Std. Error of Kurtosis
Skewness

Cell Statistics:

Number of Cases



Continue

Cancel

Help

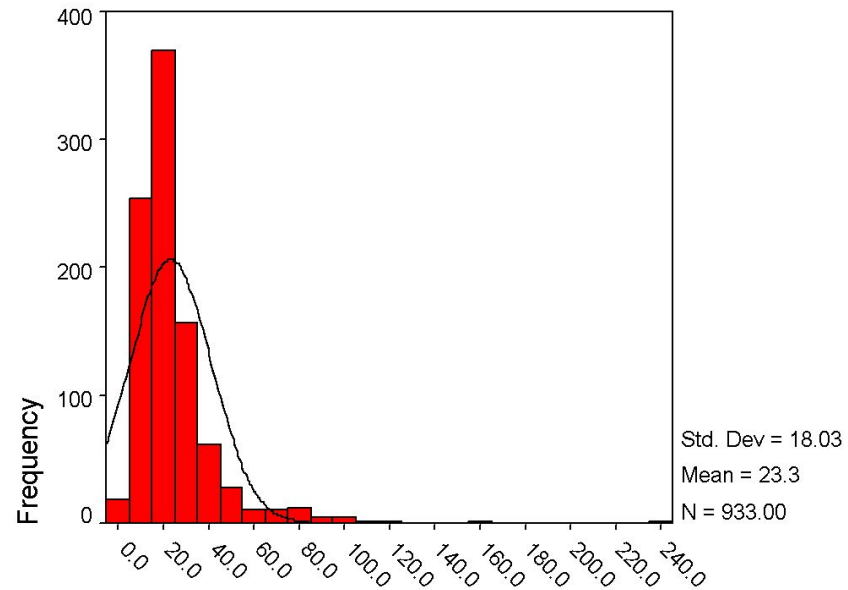
TABLE 1

Variable	EXPIRED	Survived	Total
Frequency	870	835	1705
Percent	50.5	48.5	100.0

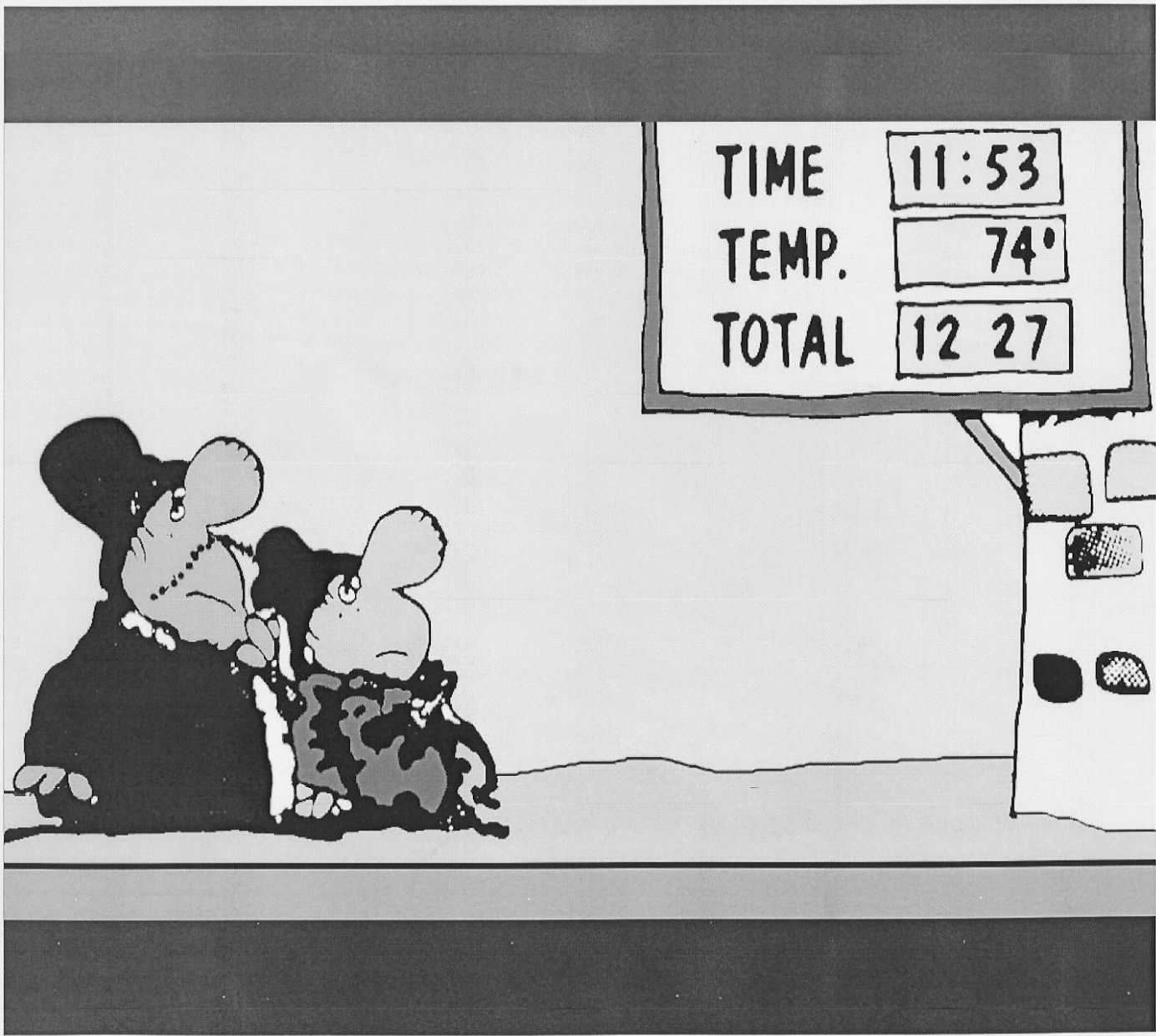
Statistics

Statistic	Value
N	933
Mean	23.34
Median	10.00
Mode	50
Std. Deviation	18.03
Range	530
Minimum	1
Maximum	531

49-DAYS IN HOSPITAL



49-DAYS IN HOSPITAL



TIME

11:53

TEMP.

74°

TOTAL

12 27