И.М.Михалевич
С.П.Примина

Применение математических методов при анализе геологической информации
(с использованием EXCEL)

ИРКУТСК 2001

Министерство образования Российской Федерации

ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

И.М. Михалевич

С.П. Примина

Применение математических методов при анализе геологической информации
(с использованием EXCEL)

Учебное пособие

Часть 1

Иркутск 2001

Введение

Затянувшийся вопрос о применимости математических методов в практике геологических работ давно исчерпал себя и в настояшее время геологи различных специальностей все больше стараются применять в своей повседневной работе численные методы. При этом геологи часто недостаточно представляют себе и сами математические (статистические)методы и что они могут дать геологу – исследователю в решении практических задач.

Применение же численных методов целесообразно на всех этапах геологических работ – от анализа первичной геологической информации до использования их при построении прогнозных разрезов и карт различного профиля и при подсчете запасов того или иного полезного ископаемого. При этом арсенал математических методов весьма обширен.

Первый выпуск данного пособия предназначен для ознакомления с основами теории вероятности и с началами математической статистики и включает в себя вопросы, связанные с изучением распределения вероятностей, теорией выборочного метода, проверкой статистических гипотез, анализа вариационного ряда, ознакомлением с некоторыми критериями сравнения выборок.

Учебное пособие состоит из четырех частей, первые три части которого знакомят с теоретическими вопросами, с наглядными примерами. Четвертая часть посвящена использованию электронной таблицы EXCEL при статистических расчетах и предназначена для на практических занятий по курсам, связанным как с изучением численных методов в геологии, так и с изучением самой электронной таблицы.

Часть 1. ВЕРОЯТНОСТЬ

В любом явлении существует некоторое множество (иногда бесконечное) возможных исходов, которые находятся в вероятностной зависимости, описываемой частотой встречаемости. Анализируя вероятности этих исходов, можно оценить поведение изучаемого объекта или события в прошлом или будущем.

Каждый из нас имеет интуитивное представление о вероятности. Например, если вас спросят, будет ли завтра дождь, вы ответите с некоторой степенью уверенности, что дождь либо возможен, либо невозможен, или в редком случае, что дождь будет обязательно, или, наоборот, что дождя наверняка не будет. Одним из способов выражения нашей уверенности является числовая шкала, например, процентная. Если мы полагаем, что вероятность того, что пойдет дождь 30%, то вероятность того, что дождя не будет, равна 70%.

Обычно выражают вероятность числами от 0 до 1 или эквивалентным числом процентов от 0 до 100%. Если мы говорим, что вероятность дождя завтра 0, мы абсолютно уверены в том, что дождя завтра не будет. Однако, если мы говорим, что вероятность дождя 1, мы абсолютно уверены в том, что дождь будет. Вероятность, выраженная таким способом, соответствует утверждению о правдоподобии события. Абсолютной уверенности соответствуют крайние значения шкалы, а различной степени уверенности — промежуточные. Например, если мы говорим, что вероятность дождя завтра равна 1/2 (а значит, и того, что дождя не будет, тоже 1/2), мы выражаем свою точку зрения с наименьшей степенью уверенности. Если мы говорим, что вероятность дождя равна 3/4 (а того, что дождя не будет, 1/4), мы выражаем меньшую степень неуверенности, так как вероятность дождя в 3 раза больше вероятности того, что дождя не будет.

Наши оценки вероятности появления дождя могут быть основаны на многих факторах, а также на субъективных ощущениях. Мы будем использовать подход, позволяющий на основании предшествующего поведения явления, (например, погода), предвидеть его будущее. Такой подход к определению вероятности через «относительные частоты» интуитивно приемлем, так как эта концепция тесно связана с понятием однородности. В некоторых случаях бывают полезны другие подходы к определению вероятности, но мы не будем на них останавливаться, так как они подробно изложены в работах /1,2/.

Пример с дождем является дискретным, так как здесь дождь может быть, а может и не быть. Классический дискретный пример, приводимый во всех руководствах по теории вероятностей, связан с бросанием правильной монеты. Одно бросание имеет два исхода: герб и решка, которые равновероятны. Поэтому вероятность выпадения герба равна 1/2. Это, конечно, означает не то, что в каждом втором бросании выпадет герб, а скорее то, что при достаточно большом числе бросаний приблизительно половина исходов — гербы. Бросание монеты — очень хороший пример для иллюстрации определения вероятности. Событие имеет два исхода, и один из них обязательно будет иметь место; если не учитывать очень малую вероятность того, что монета может встать ребром, то всегда выпадет либо герб, либо решка.

На основе схемы бросания монеты можно построить интересную последовательность вероятностей. Если вероятность выпадения герба равна 1/2, то вероятность выпадения двух гербов при двух бросаниях равна $1/2*1/2= 1/4$. Далее вероятность выпадения трех гербов при трех бросаниях равна $1/2 \cdot 1/2 \cdot 1/2 = 1/8$. Обоснование этой прогрессии простое. В первом бросании вероятность выпадения герба 1/2. Так как исход второго бросания не зависит от первого, наши возможности получить герб во втором бросании снова 1/2. Подобно этому, третье бросание не зависит от предыдущих, и вероятность выпадения герба в третьем бросании снова равна 1/2. Итак, вероятность выпадения трех гербов составляет «половину половины половины».

Предположим, что теперь нам нужно определить вероятность выпадения только одного герба в трех бросаниях.

Имеются следующие возможные исходы (Г—герб; Р—решка):

ГГГ ГРГ РРР
ГГР РГГ [РГР]
[ГРР] [РРГ]

В скобках взяты те комбинации, которые удовлетворяют нашим требованиям, т. е. они содержат только один герб. Так как имеется 8 различных комбинаций, вероятность получения только одного герба при трех бросаниях равна 3/8.

Полученное нами число представляет собой число возможных комбинаций из трех по одному. Обобщая этот пример, можно определить число возможных сочетаний из n элементов по r.

Это число символически изображается так $\left(\frac{n}{r}\right)$

Можно доказать, что число сочетаний из n элементов по r равно

$$\frac{n!}{r!(n-r)!}$$

Восклицательный знак обозначает факториал, и n! есть произведение n последовательных целых чисел:

n!==1,2,3,...,n. Значение 3! равно 3*2*1=6. В нашей задаче

$$\left(\frac{3}{1}\right) = \frac{3!}{1!(3-1)!} = \frac{3*2*1}{1*(2*1)} = \frac{6}{2} = 3,$$

т. е. имеется три возможные комбинации, содержащие один герб. Так же вычислим число возможных комбинаций, содержащих два герба:

$$\left(\frac{3}{2}\right) = \frac{3!}{2!(3-2)!} = \frac{3*2*1}{2*1(!)} = \frac{6}{2} = 3$$

ГГГ [ГРГ] РРР
[ГГР] [РГГ] РГР
ГРР РРГ

Требуемые комбинации взяты в скобки. Наконец, сколько возможных комбинаций при трех бросаниях содержат три герба?

$$\left(\frac{3}{3}\right) = \frac{3!}{3!(3-3)!} = \frac{3*2*1}{3*2*1(1)} = 1.$$

Заметим, что по определению число 0! равно не нулю, а 1. Наконец, остается возможность, когда при трех бросаниях не выпадает ни одного герба, и число таких комбинаций будет равно

$$\left(\frac{3}{0}\right) = \frac{3!}{0!(3-0)!} = \frac{3*2*1}{1(3*2*1)} = 1$$

Таким образом, при трех бросаниях монеты имеется одна возможность не получить ни одного герба, три возможности получить один герб, три возможности получить два герба и одна возможность получить три герба (рис. 1.1.)

Можно подсчитать общее число возможных исходов, которое равно восьми, а затем преобразовать частоты появления отдельных событий в вероятности. Иначе говоря, вероятность отсутствия гербов при трех бросаниях, т. е. вероятность выпадения одной комбинации среди восьми возможных, равна 1/8. Преобразуем теперь гистограмму(рис. 1.1) , откладывая по вертикальной оси вместо числа комбинаций вероятности соответствующих событий, получим дискретное распределение вероятностей, график которого представлен на рис. 1.2. Общая сумма всех вероятностей равна 8/8 или 1. Таким образом, событие, заключающееся в том, что какая-либо комбинация при трех бросаниях осуществится, является достоверным.

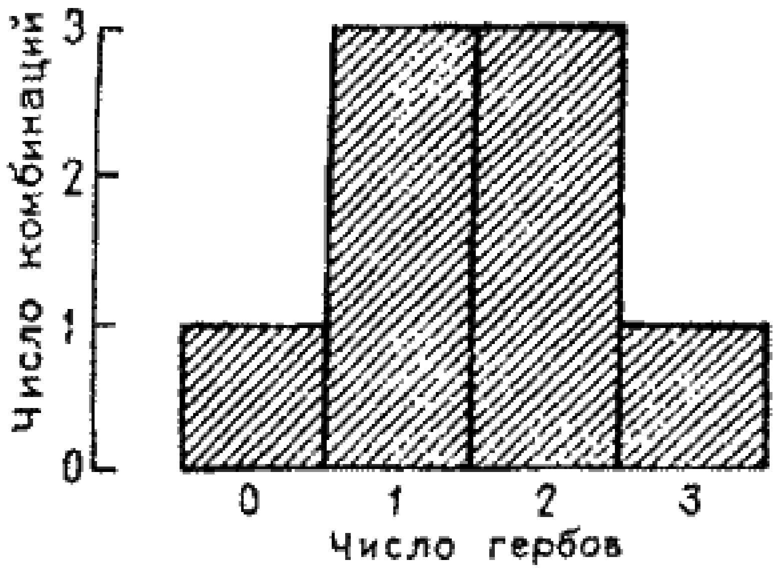


Рис 1.1

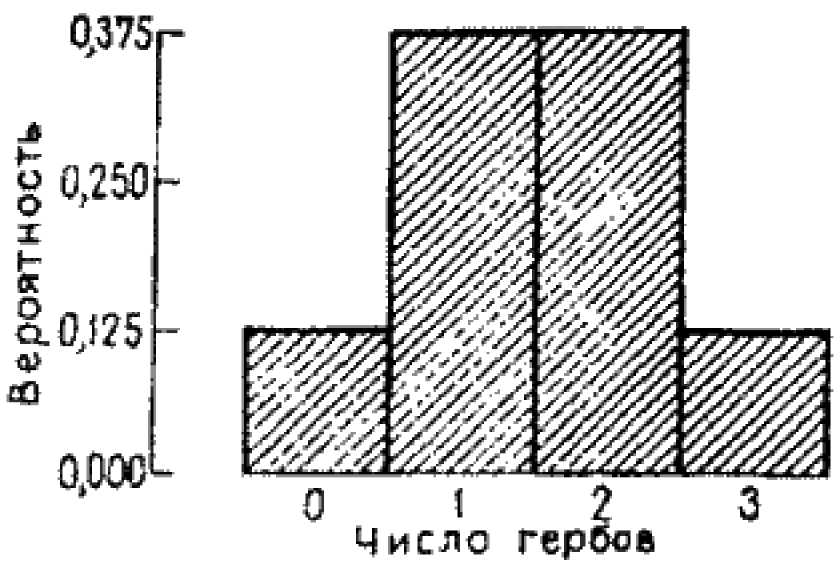


Рис 1.2

Рис. 1.1. Число комбинаций. Заштрихованная область на графике показывает число различных способов получения заданного числа гербов при трех бросаниях монеты
Рис. 1.2. Дискретное распределение, показывающее вероятность получения заданного числа гербов при трех бросаниях монеты

График функции (рис.1.2)распределения характеризует вероятность выпадения какой-либо заданной комбинации.

Эксперимент бросания монеты обладает четырьмя особенностями:

1) в каждом испытании (или бросании) имеется только два возможных исхода (назовем их «успех» или «неудача»);

2) исход каждого испытания не зависит от предыдущих исходов;

3) вероятность успеха не меняется от испытания к испытанию;

4) испытания повторяются заданное число раз.

Распределение вероятностей, соответствующее указанному нами типу эксперимента, называется биномиальным распределением.

В качестве примера геологических приложений можно упомянуть предсказание вероятности успеха при бурении нефтяных и газовых скважин. Четыре перечисленных выше характеристики должны предполагаться справедливыми: допущения подобного рода кажутся наиболее приемлемыми при использовании метода «дикой кошки» при разведке в сравнительно мало изученном бассейне. Следовательно, биномиальное распределение часто используется для предсказания результатов бурения на границе области или за пределами концессии.

В предположении о справедливости биномиального распределения каждый результат применения метода «дикой кошки» может быть классифицирован либо как «открытие» («успех») либо как пустая скважина («неудача»). Последовательные испытания предполагаются независимыми, т. е. успех или неудача в одной скважине не влияет на результат опробования в другой скважине. (Это допущение трудно обосновать по многим обстоятельствам, так как открытие обычно влияет на результаты исследования следующих скважин. Появление последовательности пустых скважин также приводит к изменению программы разведки.) Наконец, биномиальное распределение оказывается подходящим, если фиксированное число скважин будет пробурено в процессе разведки или в течение единичного периода времени (например, бюджетного цикла), для которого проведено прогнозирование.

За вероятность p открытия нефтяного или газового месторождения методом «дикой кошки» принимаются широко используемые в промышленности отношения, характеризующие успех бурения, которые были наблюдены в процессы бурения в аналогичных регионах, или аналогичные отношения, полученные компаниями, делающими оценки, или же просто субъективные аналогичные характеристики. Для определения вероятности p в биномиальной модели в разведочном бурении предусмотрены следующие этапы:

1) вероятность «успеха» бурения скважины обозначим через p;

2) вероятность «неудачи» бурения обозначим через 1—p;

3) вероятность того, что в результате последовательного бурения n скважин методом «дикой кошки» все окажутся пустыми, есть

$$P = (1 - p)^{n}.$$

4) вероятность того, что n-я пробуренная скважина окажется продуктивной, причем n - 1 предыдущих скважин окажутся пустыми, равна

$$P = (1 - p)^{n-1} p ;$$

5) вероятность появления одной продуктивной скважины в последовательности n скважин, пробуренных методом «дикой кошки», есть

$$P = n (1 - p)^{n-1} p;$$

так как продуктивной может оказаться любая из n скважин;

6) вероятность того, что будет пробурено n-r пустых скважин, за которыми последует r продуктивных, есть

$$P = (1 - p)^{n-r} p^{r};$$

7) однако n - r пустых скважин и r продуктивных скважин могут быть выбраны $\binom{n}{r}$ способами, или, что эквивалентно, $\frac{n!}{(n-r)!r!}$ различными способами. Так, вероятность того, что в программе бурения методом «дикой кошки» среди n скважин получим r продуктивных, равна

$$P = \frac{n!}{(n-r)!r!} (1 - p)^{n-r} p^{r}. \qquad (1.1)$$

Это - формула биномиального распределения, задающая вероятность получения r успехов в последовательности n независимых испытаний, вероятность успеха в каждом из которых равна p.

Уравнение (1.1), которое может быть решено, т. е. может быть определена вероятность появления любой заданной комбинации успехов и неудач при любом заданном числе испытаний и заданной вероятности.

Например, мы хотим определить вероятности, соответствующие программе разведки в малоизученном бассейне, содержащем пять скважин, причем отношение, характеризующее успех. в этом регионе, равно примерно 10%. Какова вероятность того, что вся программа будет полностью безуспешной и мы не получим ни одной продуктивной скважины? Такой результат называется полным крахом по очевидным причинам, и биномиальное выражение содержит следующие члены

$$n = 5;\ r = 0;\ p = 0.10;$$

$$P = \binom{5}{0} * 0.10^{0} * (1 - 0.10)^{5} = \frac{5!}{5!0!} * 1 * 0.90^{5} = 1*1*0.59 = 0.59.$$

Вероятность того, что в результате разведки не будет получено никаких продуктивных скважин, равна почти 60%.

(Пример расчета вероятности с помощью MS Excel на рис 4.1, часть 4 пособия).

Если только одна скважина окажется продуктивной, то она может окупить все затраты в процессе разведки. Какова вероятность того, что одна скважина окажется продуктивной в процессе разведки с пятью скважинами?

$$P = \binom{5}{1} * 0.10^{1} * (1 - 0.10)^{4} = \frac{5!}{4!1!} * 0.10 * 0.90^{4} = 5*0.10*0.656 = 0.328.$$

Перед тем как перейти к следующей теме , следует вернуться к биномиальному распределению. На рис.1.2 представлено вероятностное распределение для всех возможных чисел выпадения гербов и решек при трех бросаниях монеты. Аналогичный эксперимент можно осуществить при большем числе испытаний. На рис 1.3, например, представлены вероятности получения заданного числа «успехов» (гербов) в десяти бросаниях монеты, а на рис. 1.4 - вероятностное распределение, которое описывает результаты эксперимента, состоящего из 50 бросаний монеты. Все эти вероятности были получены из таблиц биномиального распределения ,или могут быть вычислены из биномиального уравнения или с помощью электронной таблицы EXCEL , или с помощью статистических прикладных пакетов программ .

В каждом из этих экспериментов вычислялись все возможные числа гербов, которые можно получить, начиная с 0 до 3, 10, 50. Никакие другие комбинации гербов и решек не могут встретиться.

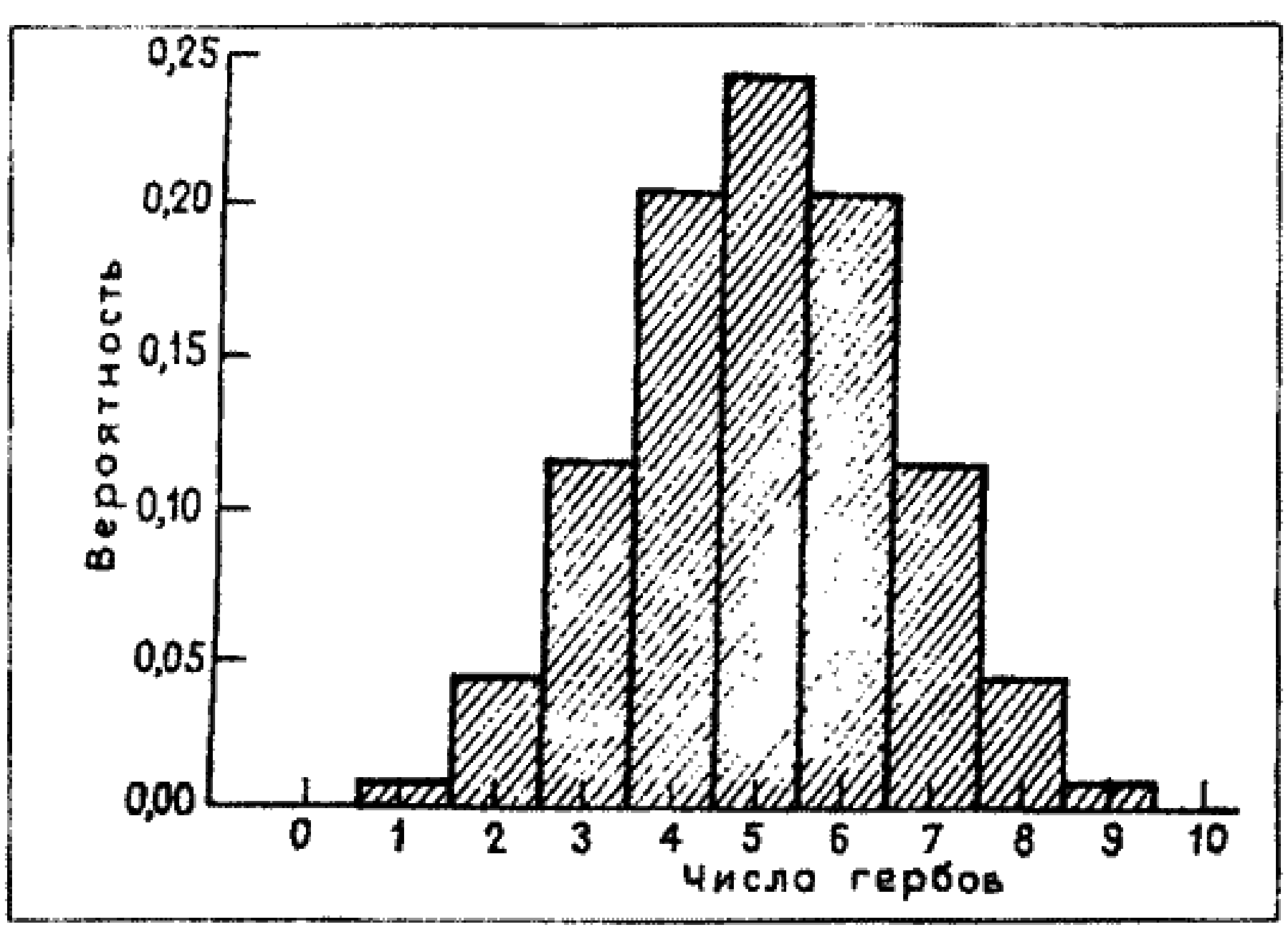


Рис 1.3. Дискретное распределение , показывающее вероятность заданного числа гербов при 10 бросаниях монеты

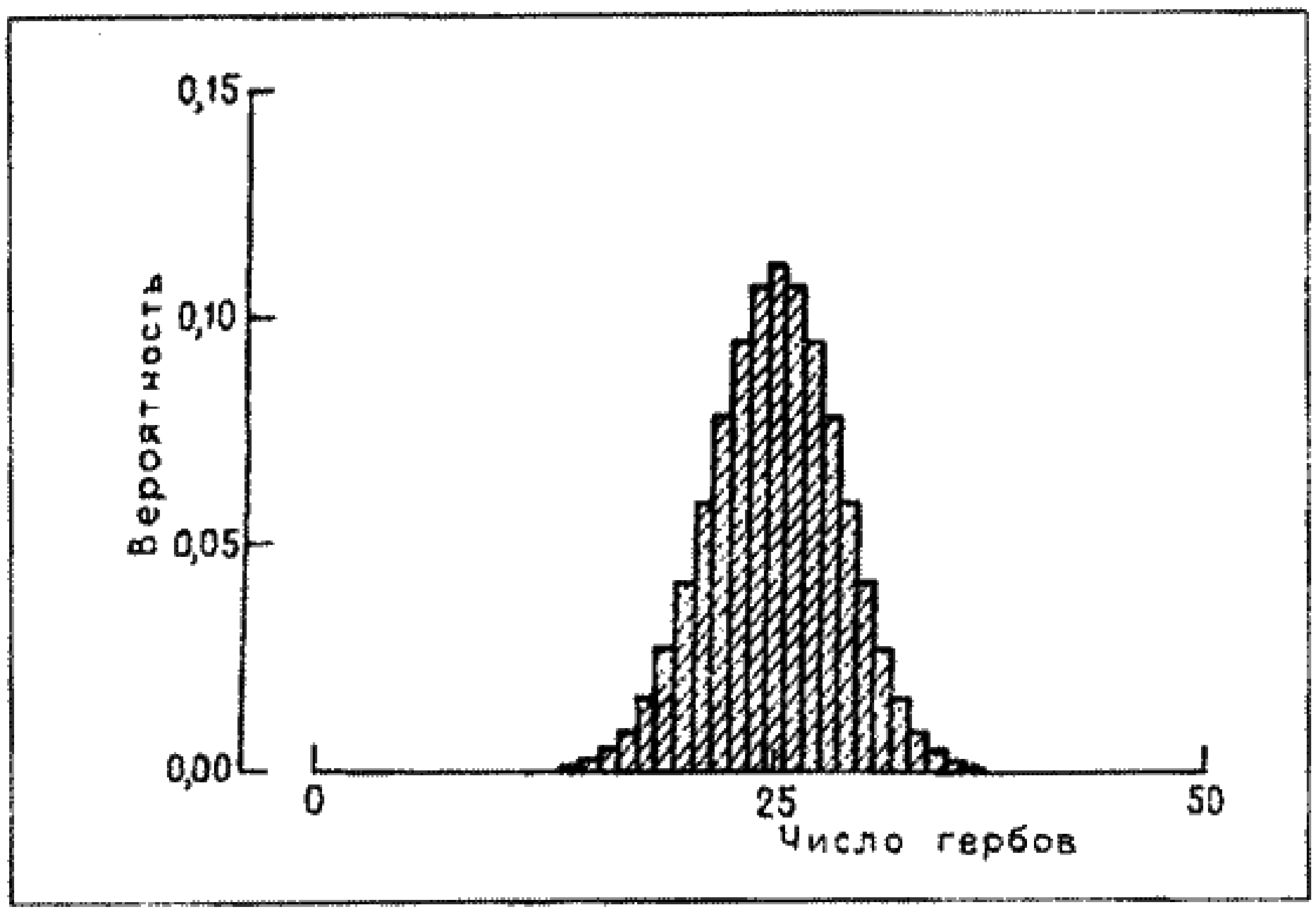


Рис 1.4. Дискретное распределение , показывающее вероятность заданного числа гербов при 50 бросаниях монеты

Так как мы обязательно получаем какой-либо результат из перечисленных выше, то сумма всех вероятностей в каждом эксперименте должна равняться 1,00. Это удобно представить, полагая площади под гистограммами на рис. 1.3 и 1.4 равными 1,00, как это сделано на гистограмме, изображенной на рис. 1.2. При таком условии увеличивающееся число бросаний монеты будет сопровождаться только сужением ширины полос.

Гистограмма становится все более напоминающей гладкую и непрерывную кривую. Можно представить себе эксперимент, состоящий в бесконечном числе бросаний монеты, в результате которого будет получено бесконечное число полос бесконечно малой ширины. Тогда гистограмма превратится в непрерывную кривую, и горизонтальная ось будет представлять скорее непрерывную, чем дискретную переменную.

В эксперименте бросания монеты мы имеем дело с дискретными исходами, т. е. со специфической комбинацией гербов и решек. Однако в большинстве экспериментов, встречающихся на практике, возможные исходы не являются дискретными. Обычно имеется бесконечный континуум возможных исходов, которые могут быть получены. Множество возможных исходов может быть конечным и на самом деле ограниченным, но в пределах этого множества результат, который может быть получен, нельзя предсказать точно. В данном случае мы имеем дело с непрерывными случайными переменными. Предположим, например, что измеряется рост человека и установлено, что он равен 176 см. Однако если провести измерение, используя более точный способ, то можно получить рост 176,2 см, а при использовании еще более точного инструмента измерения - 176,23 см, и т.д.

Непрерывная переменная теоретически может бесконечно подразделяться. Это является следствием того факта, что всегда можно найти разность между двумя измерениями, если проводить измерения в достаточно мелкой шкале. Следствие этого утверждения таково, что каждый исход в непрерывной шкале измерений уникален и что вероятность получения конкретного точного результата равна нулю.

Если это так, то кажется невозможным определить вероятность на основе относительных частот встречаемости. Однако, хотя невозможно наблюдать исходы, которые в точности соответствуют 176,000...000 см, вполне доступно получить ряд измерений, попадающих внутрь некоторого интервала, включающего это значение. При том, что индивидуальные изменения в точности не идентичны, они достаточно близки и можно считать их принадлежащими одному классу. В итоге разобьем непрерывную шкалу на дискретные сегменты, и тогда можно подсчитать число событий, попавших внутрь каждого интервала. Сужая границы класса, мы уменьшим число событий в нем, и снизим оценки вероятностей появления события.

Если мы имеем дело с дискретными событиями, то определяем значения с абсолютной точностью. Непрерывные переменные, однако, измеряются с помощью некоторых физических процедур, которые ограничивают точность их измерения. В повторных измерениях, сделанных на одном и том же объекте, возникают малые отклонения, величина которых отражает как естественные изменения объекта, так и изменения в условиях проведения измерений и, кроме того, изменения, обусловленные деятельностью исследователя, производящего измерения. Единственное, точное, «истинное» значение не может быть определено; иными словами, мы наблюдаем непрерывное распределение возможных значений. Такие свойства присущи непрерывной случайной (или стохастической) переменной (величине) .

Для иллюстрации непрерывных случайных величин рассмотрим задачу определения показателя проницаемости образцов из керна скважины. (Этот пример достаточно показателен для различных видов изменчивости непрерывных случайных величин.) Керн – цилиндрический столбик горной породы, остающийся внутри бурового снаряда при бурении и являющийся важным геологическим документом для изучения.

Проницаемость определяется временем, требуемым для проникновения заданного количества флюида при стандартных условиях через образец породы. Допустим, что в результате одного определения получена проницаемость, равная 0,108 единицам измерения проницаемости. Является ли это число «истинной» проницаемостью пробы? Другие определения на этом же образце могут дать проницаемость, равную 0,093 и 0,112 ед. На проницаемость, записываемую приборами в ходе любого эксперимента, влияет ряд условий, которые внутри прибора неизбежно изменяются от одного определения к другому и не зависят от действий оператора. Ни одно из полученных значений нельзя взять в качестве абсолютной меры истинной проницаемости. В итоге порождается непрерывная случайная величина, которую мы подвергаем опробованию, делая повторные измерения.

Изменчивость, обусловленная неточностью инструментов, более очевидна, когда делаются повторные измерения на единичном объекте, т. е. испытания повторяются без изменений. Такую изменчивость называют ошибками эксперимента. Кроме этого, изменчивость может проявляться в последовательности измерений или результатов экспериментов, проводимых на ряде изучаемых объектов. Обычно именно эта изменчивость и представляет научный интерес. Довольно часто оба эти типа изменчивости так перепутаны или совмещены, что экспериментатор не может определить, какая часть изменчивости возникает в силу различий между условиями испытаний, а какая является следствием ошибок измерения.

Предположим, что у нас не образец породы, а значительной длины керн, взятый из скважины, проходящей через слой песчаника. Мы хотим определить проницаемость песчаника, но не можем ввести керн длиной в 20 условных единиц в аппарат для измерения проницаемости. Вместо этого мы вырежем из керна несколько малых частей (интервалов) и определим проницаемость каждого из них. Наблюдаемая изменчивость явится следствием различий как между испытуемыми частями керна, так и между условиями эксперимента.

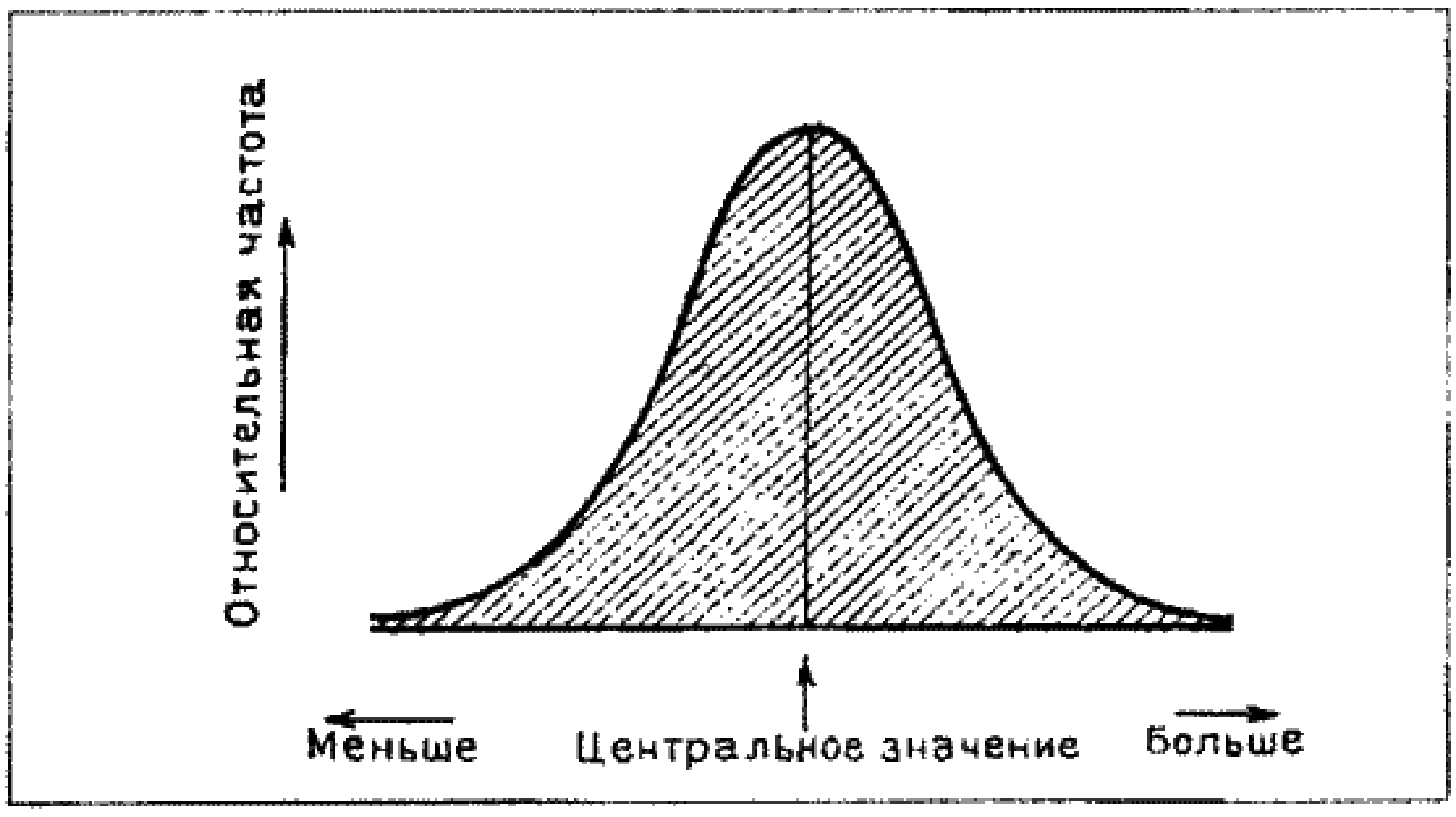


Рис. 1.5. График плотности нормального распределения

Разработка методов оценки величины отдельных источников изменчивости - одна из важнейших задач статистики.

Повторные измерения, проводимые на больших выборках, взятых из естественных совокупностей, дают возможность охарактеризовать распределение частот. Обычно большая часть значений группируется около некоторого центрального значения, при удалении от которого частоты убывают. График (рис. 1.5) имеет один максимум и называется нормальным распределением. В прикладных работах с использованием статистических методов часто делается допущение, что случайные величины распределены нормально, и многие статистические критерии основаны на этом допущении.

Общую площадь, заключенную между графиком нормального распределения и горизонтальной осью, можно считать равной 1,00 (или 100%). Поэтому, используя график, можно вычислить вероятность соответствующего события. Здесь нельзя не обратить внимание на сходство одновершинной непрерывной кривой(на рис. 1.5) с гистограммой, (на рис. 1.4).Однако поскольку в случае непрерывного распределения число подразделений по горизонтальной оси можно считать бесконечным, вероятность получения какого-либо конкретного значения равна нулю. Вместо этого мы рассмотрим вероятность появления значений в пределах некоторого заданного интервала. Эта вероятность равна площади под кривой частот, заключенной между заданными пределами. Если указанный промежуток велик, то осуществление события в этом промежутке представляется более правдоподобным. Если интервал очень мал, то появление события маловероятно.

Выше были введены без определения два важных статистических понятия - «совокупность» и «выборка». Совокупность (генеральная совокупность) состоит из вполне определенного множества (либо конечного, либо бесконечного) элементов. Так, количество марсиан на Марсе есть конечная совокупность, так как мы “знаем”,что их там всего 200 (по примеру из следующего раздела), температура же тела или же количество песчинок на Земле с замеренным их весом есть бесконечные совокупности. Вообще эти элементы можно рассматривать как измерения, выполненные на объектах заданного типа.

Выборка — это подмножество элементов, выбранных из некоторой совокупности.

Если наблюдения с заданными свойствами систематически исключаются из выборки, то такую выборку называют смещенной. Предположим, например, что нас интересует проницаемость данного слоя горной породы из керна. Если из выборки исключить все рыхлые и раздробленные породы, так как их проницаемость трудно измерить, то результат изменится. Вероятно, полученный интервал значений пористости будет усечен справа, что даст смещение выборки в сторону более низких значений, и потому мы получим ошибочно заниженную оценку изменчивости проницаемости в слое.

Обычно выборки извлекаются из совокупности наудачу. Это значит, что все элементы совокупности имеют равные возможности быть включенными в выборку. Случайная выборка будет несмещенной, и по мере возрастания ее объема она будет точнее описывать рассматриваемую совокупность. Такие выборки обычно называют представительными или репрезентативными.

Заканчивая этот раздел мы хотели бы подчеркнуть , что о вероятности нами написано самое необходимое для последующего понимания статистического анализа данных. Более подробно о вероятности можно найти в работах /1,2,3,4/.

Часть 2. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Вернемся к распределениям. Распределения имеют ряд характеристик, например, такие, как средняя точка, меры разброса и меры симметрии. Эти характеристики называются параметрами, если они описывают совокупности, и статистиками, если они относятся к выборкам. Статистики можно использовать для оценки параметров

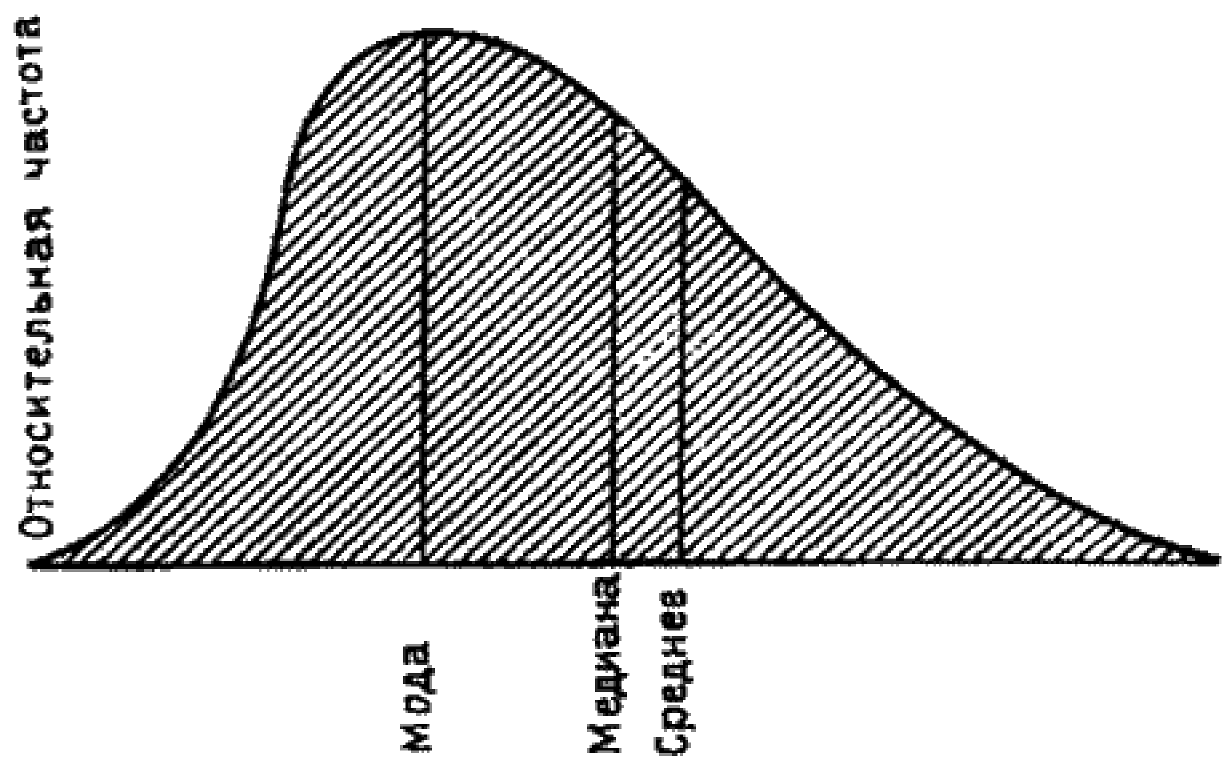


Рис. 2.1. Соотношение между видами средних значений в ассиметричном частотном распределении

исходных совокупностей и для проверки гипотез, сформулированных относительно этих совокупностей.

Наиболее очевидная характеристика совокупности или выборки—ее среднее значение. Существуют различные виды среднего значения, но только некоторые из них используются на практике. Мода — значение, которое соответствует наибольшей частоте. Например, в распределении, приведенном на рис. 2.1, мода соответствует наивысшей точке кривой частот, а медиана—средняя точка распределения. На рис. 2.1 показано, что половина площади под кривой распределения находится справа от медианы, а другая половина — слева. Среднее значение — это, иными словами, среднее арифметическое, которое определяется как сумма всех результатов наблюдений, деленная на их число. В условиях асимметричных кривых распределения медиана расположена между средним значением и модой, а в случае симметричных кривых, подобных нормальной, все три меры совпадают.

Некоторые символы традиционно используются в качестве характеристик кривых распределения. Обычно для обозначения характеристик теоретических распределений используются греческие буквы, а для выборочных —латинские. Так, например, выборочное среднее обозначается $\overline{X}$, а теоретическое среднее значение всей совокупности μ (читается "мю"). Основная задача обычно заключается в том, чтобы оценить некоторые параметры изучаемого распределения. Статистика, которую мы вычисляем по выборке из взятой совокупности, используется как оценка требуемого параметра. Применение греческих и латинских букв подчеркивает разницу между параметрами и соответствующими им статистиками.

Среднее арифметическое, вычисленное по данным выборки, имеет два в высшей степени желательных свойства, которые делают его более полезным для оценки среднего или центрального значения распределения, чем любая из двух других выборочных характеристик: медиана или мода. Во-первых, среднее арифметическое является несмещенной оценкой истинного среднего значения совокупности. Необходимо отметить, что статистика — это несмещенная оценка соответствующего параметра, если ее среднее значение, взятое по большому набору выборок, равно этому параметру. Во-вторых, можно показать, что для симметричных распределений, подобных нормальному, среднее арифметическое характеризуется тенденцией лучшего приближения к среднему значению совокупности, чем любая другая несмещенная оценка (такая, как медиана), построенная по той же выборке. Это равносильно тому, что выборочные средние имеют меньший разброс, чем выборочные медианы, и, следовательно, являются более эффективными.

Другая характеристика распределения — мера разброса отдельных значений относительно среднего, или дисперсия. Известны различные меры этого свойства, но только две из них широко используются. Одна из них — уже упомянутая дисперсия, а другая — квадратный корень из дисперсии, называемый стандартным отклонением. Дисперсию можно рассматривать как среднее значение квадратов отклонений всех возможных значений случайной величины от истинного среднего совокупности.

Рассмотрим эффектный пример из работы Гланца /6/. По Гланцу , экспедиция слетала на Марс и измерила всех марсиан, благо их всего две сотни. Результаты приведены на рис. 2.2 (рост округлен до целого числа сантиметров). Каждому марсианину соответствует кружок, так что, например, два кружка над числом 30 означают, что имеются два марсианина ростом 30 см. Рис. 2.2 —это распределение марсиан по росту. Мы видим, что рост большинства марсиан—от 35 до 45 см. Коротышек (ниже 30 см) совсем немного — всего трое, и столько же великанов (выше 50 см).

Слетаем на Венеру и измерим всех 150 ее обитателей. Отчет об экспедиции будет звучать так: «Редко встретишь венерианца ниже 10 см или выше 20 см, а чаще попадаются 15-сантиметровые, см. рис. 2.3».

Но вот остались позади межпланетные перелеты. Настала пора анализа данных. Сравним рис. 2.2 и 2.3. Мы видим, что венерианцы ниже марсиан и что интервал,

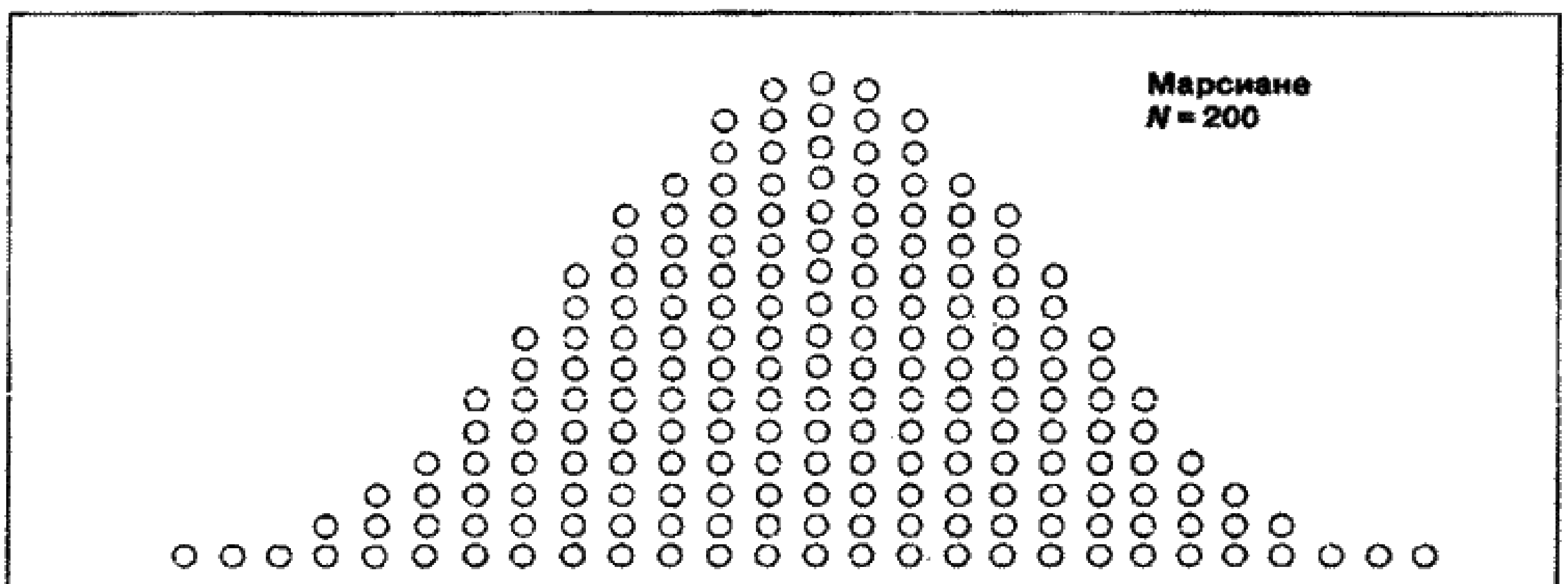


Рис. 2.2. Распределение марсиан по росту. Каждому марсианину соответствует кружок. Обратите внимание, что марсиан среднего роста (около 40 см) больше всего и что высокорослых столько же, сколько коротышек (распределение симметрично)

в который умещается рост всех марсиан, шире, чем соответствующий интервал для венерианцев. Ширина интервала, в который попадают почти все марсиане (194 из 200) — 20 см (от 30 до 50 см). Рост большинства венерианцев (144 из 150) умещается в интервал от 10 до 20 см, то есть имеет ширину всего лишь 10 см. Несмотря на эти различия, между двумя совокупностями инопланетян имеется и существенное сходство. В обеих рост любого члена скорее близок к середине распределения, нежели заметно от нее удален, и одинаково вероятно может быть как выше, так и ниже середины. Распределения на рис. 2.2 и 2.3 имеют схожую форму и приближенно определяются одной и той же формулой.

Раз существует множество похожих распределений, значит, для характеристики одного из них достаточно указать, чем оно отличается от других, ему подобных, то есть всю собранную информацию мы можем свести к нескольким числам, которые называются параметрами распределения.

Это,в первую очередь,среднее значение и стандартное отклонение

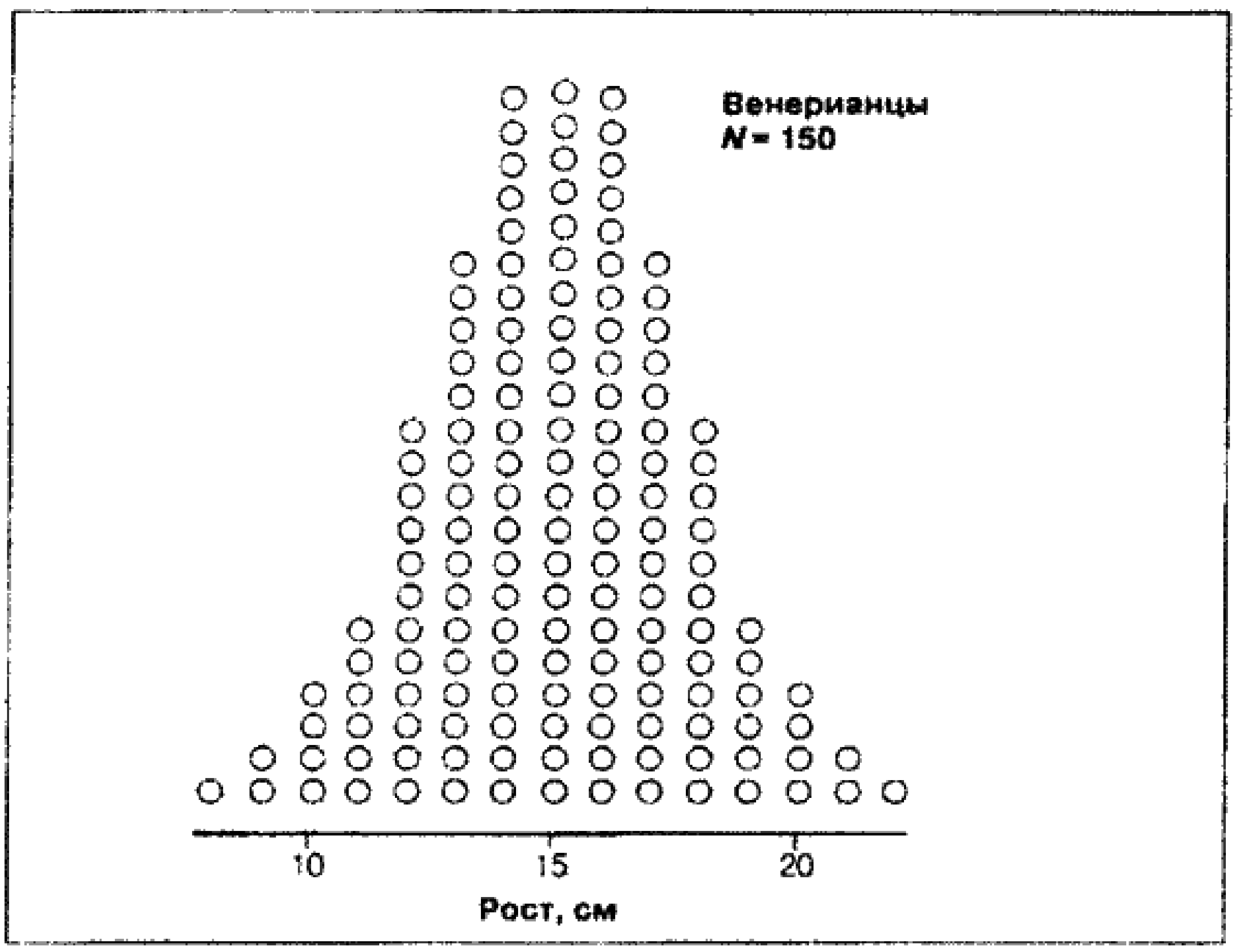


Рис. 2.3 Распределение венерианцев по росту.(Венерианцы ниже марсиан, разброс значений меньше. Однако по форме распределения, напоминающей колокол, венерианцы и марсиане схожи друг с другом.)

Расположив мысленно распределения марсиан и венерианцев на одной шкале роста, мы увидим, что распределение венерианцев находится ниже, чем распределение марсиан. Характеристика положения распределения на числовой оси называется средним. Среднее по совокупности вычисляют по формуле:

$$\text{Среднее по совокупности} = \frac{\text{Сумма значений признака для всех членов совокупности}}{\text{Число членов совокупности}}$$

Эквивалентное математическое выражение имеет вид

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} . \qquad (2.1)$$

где X — значение признака, N — число членов совокупности. Как всегда, большая греческая буква $\sum$ (читается «сигма») обозначает сумму. Подставив в формулу добытые нами данные, получим ценное дополнение к научному отчету: средний рост марсиан 40 см, а венерианцев — 15 см. Запомним эти параметры.

Дисперсию можно рассматривать как среднее значение квадратов отклонений всех возможных значений случайной величины от истинного среднего совокупности , которая определяется по формуле

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2 \ . \qquad (2.2)$$

Этим равенством определяется истинная дисперсия совокупности σ^2 (маленькая греческая буква "сигма").

Выборочная дисперсия определяется символом s^2. Если наблюдения $X_1,\ldots,X_n$ $(n < N)$ случайная выборка из совокупности с нормальным распределением, то s^2 является оценкой для σ^2.

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n-1} \left(X_i - \bar{X} \right)^2 \ . \qquad (2.3)$$

Так как дисперсия является средним значением квадратов отклонений от среднего, то ее размерность характеризуется квадратами единиц, которыми измерялись исходные наблюдения. Как видно из формулы (2.2), дисперсия измеряется в единицах, равных квадрату единицы измерения соответствующей величины. Например, дисперсия измеряемого в сантиметрах роста сама измеряется в квадратных сантиметрах. Это довольно неудобно. Поэтому чаще используют квадратный корень из дисперсии — стандартное отклонение σ.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}} \ . \qquad (2.4)$$

Стандартное отклонение измеряется в тех же единицах, что исходные данные. Например, стандартное отклонение роста марсиан составляет 5 см, а венерианцев — 2,5 см. Запомним эти параметры.

По аналогии с выборочной дисперсией s^2 выборочное стандартное отклонение обозначается через s:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2}{n-1}} \qquad (2.5)$$

В формулах (2.3) и (2.5) расчет оценок s^2 и s проводится делением центрированной суммы квадратов на (n-1) потому ,что разброс значений в пределах выборки никогда не бывает столь большим, как во всей совокупности, и деление не на n , а на (n-1) компенсирует возникающее занижение оценки стандартного отклонения /6/.

Малое значение стандартного отклонения указывает, что наблюдения хорошо группируются около центрального значения. Наоборот, большое стандартное отклонение показывает, что наблюдения широко рассеяны относительно среднего значения и имеют слабую тенденцию к централизации.

Весьма полезное свойство нормального распределения состоит в том, что площадь под кривой в пределах некоторого заданного интервала может быть точно вычислена. Например, более 2/3 наблюдений (68,27%) попадают в интервал с центром в среднем значении и длиной, равной двум стандартным отклонениям. Примерно 95% всех наблюдений заключается в интервале от —2 до +2 стандартных отклонений и более 99% содержится в интервале от —3 до +3 стандартных отклонений на рис. 2.4.

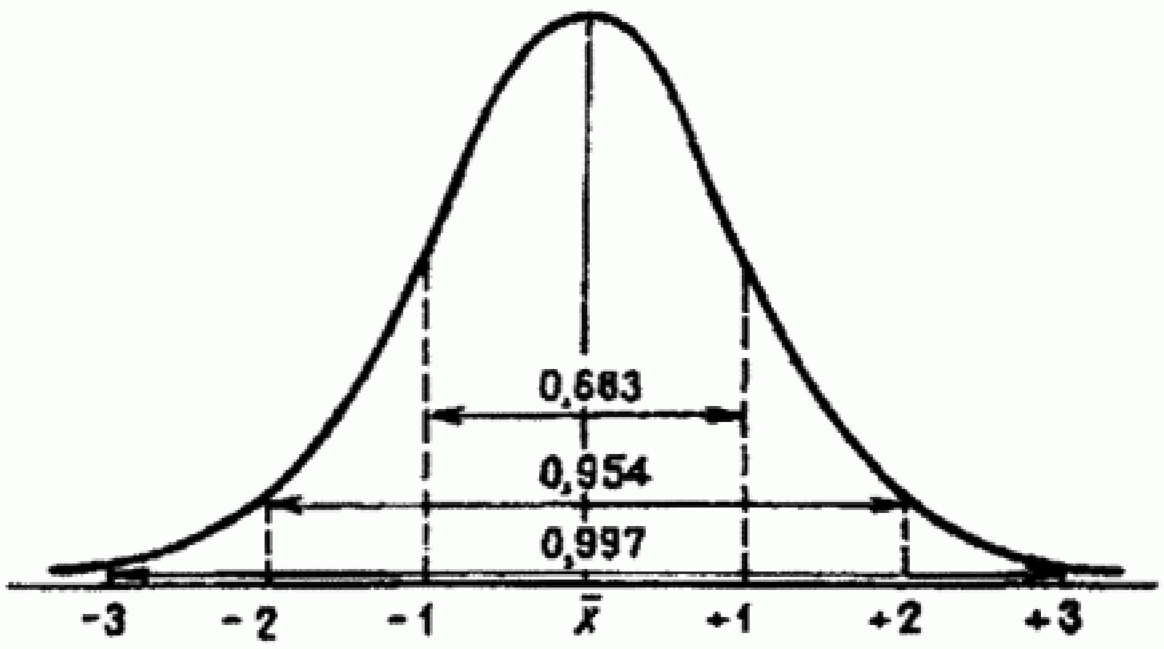


Рис. 2.4. Площади стандартного нормального распределения, заключенные в пределах интервалов, кратных стандартному отклонению

Обратившись к рис. 2.2 и 2.3, мы обнаружим, что на обеих планетах рост примерно 68% обитателей отличается от среднего не более чем на одно стандартное отклонение и примерно 95% — на два стандартных отклонения и.т.д. Эти данные подтверждают выводы проиллюстрированные на рис 2.4.

Подобные (нормальные) распределения встречаются очень часто. Можно сказать, что это происходит всегда, когда некая величина отклоняется от средней под действием множества слабых, независимых друг от друга факторов. Нормальное распределение еще называется "гауссовым" и описывается формулой :

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}, \qquad (2.6)$$

где π — постоянная величина, представляющая собой отношение длины окружности к ее диаметру (π=3,14...). Экспоненциальная форма, которую можно записать как e^x вместо exp (x), является степенью иррационального числа e. С точностью до пятого десятичного знака e = 2,71828.

Таким образом, если для заданного множества данных известны μ и σ, обозначающие их среднюю и стандартное отклонение, то их можно подставить в формулу (2.6), и значение f(X) (высота кривой) будет известно для любого заданного значения переменной X.

Нормальное распределение полностью определяется параметрами μ и σ ,поэтому сведения в табл.2.1 - не просто представление данных , но и полное их описание.

Таблица 2.1. Параметры распределения марсиан и венерианцев по росту

	Объем совокупности	Среднее, см	Стандартное отклонение, см
Марсиане	200	40	5,0
Венерианцы	150	15	2,5

Те, кто не имел пока дела со статистическим анализом, обычно с трудом интуитивно воспринимают численные значения дисперсии или стандартного отклонения. Является ли дисперсия, равная 10, большой или малой? Что значит стандартное отклонение 23? Оказывается, для интерпретации как дисперсии, так и стандартного отклонения не требуется приписывать каждому из них численного значения, а требуется сравнивать одну дисперсию с другой. Выборка, имеющая наибольшую дисперсию или стандартное отклонение, характеризуется большим разбросом наблюдаемых значений при условии, что все измерения сделаны в одних и тех единицах (см. рис 2.2, 2.3).

Теперь , имея некоторый опыт по описанию данных , вернемся к уже знакомому термину "медиана" и познакомимся с так называемыми "процентилями".

Обратимся к " галактическому примеру " С. Гланца . На Юпитере не только измерены все до одного юпитерианина, но также подсчитаны среднее и стандартное отклонение роста для всей их совокупности. Оказывается, средний рост юпитериан — 37,6 см, а его стандартное отклонение — 4,5 см. Можно заключить, что юпитериане очень похожи на марсиан, ведь близки оба параметра, определяющие нормальное распределение — среднее и стандартное отклонение.

Однако, если взглянуть на исходные данные по юпитерианам (рис. 2. 5А), то обнаружится совершенно иная картина. На самом деле типичный юпитерианин довольно приземист — около 35 см, то есть на добрых 5 см ниже марсианина. И только небольшая группа "долговязых "смещает значения стандартного отклонения и среднего, вводя исследователей в заблуждение!

Итак, рост произвольно выбранного юпитерианина вовсе не равновероятно может оказаться выше или ниже среднего, то есть распределение юпитериан по росту асимметрично (рис.2.1) .

В такой ситуации полагаться на среднее и стандартное отклонение нельзя. На рис. 2.5,Б изображено нормальное распределение для совокупности с теми же самыми значениями среднего и стандартного отклонения, что и на рис. 2.5А. Оно ничуть не похоже на распределение юпитериан. Таким образом, доверившись среднему и стандартному отклонению, мы получим превратное

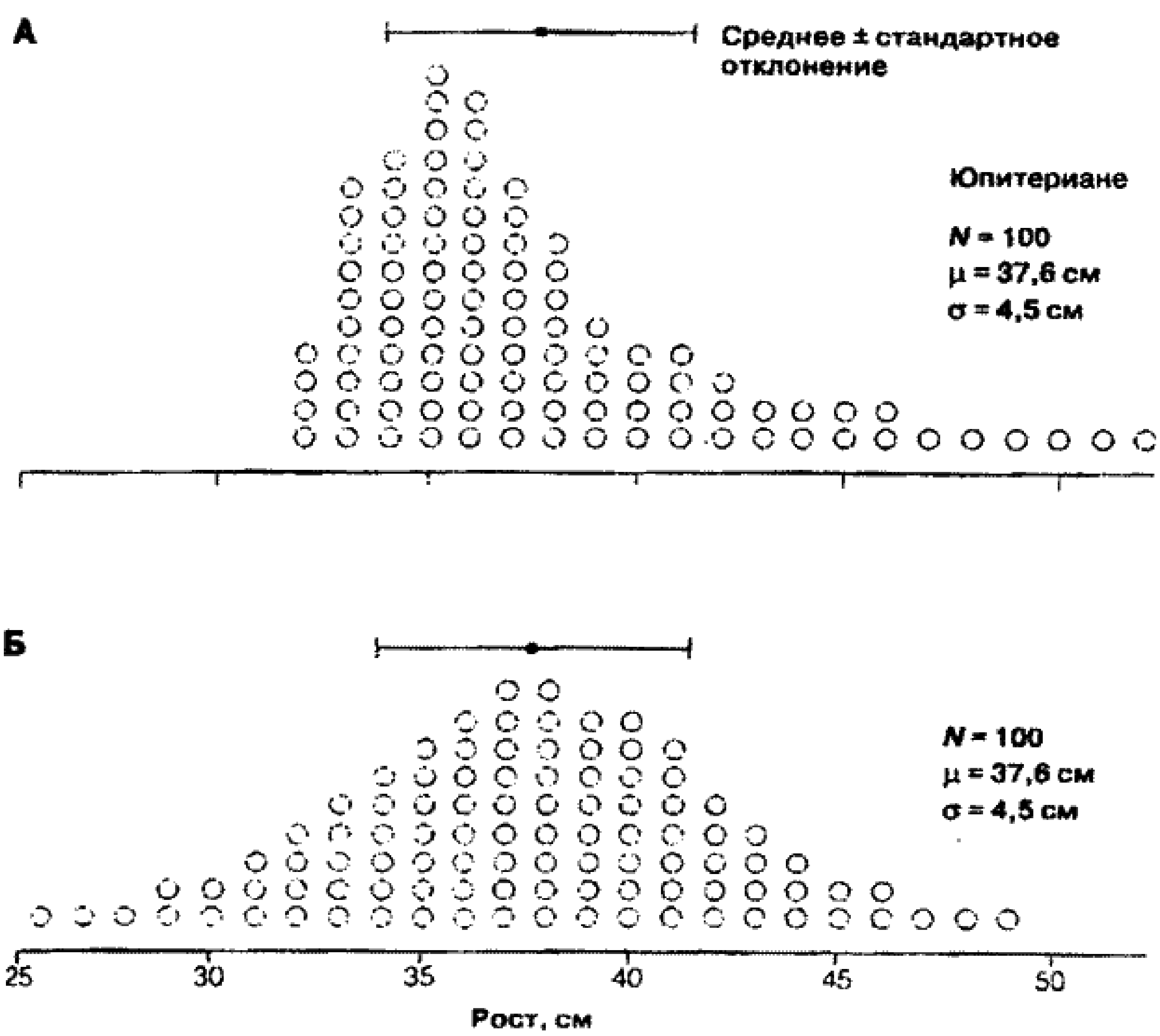


Рис. 2.5. Если распределение асимметрично, полагаться на среднее и стандартное отклонение нельзя. А. Распределение юпитериан по росту. Б. Нормальное распределение с теми же средним и стандартным отклонением: несмотря на тождественность параметров, оно ничуть не похоже на реальное распределение юпитериан

представление о совокупности, не подчиняющейся нормальному распределению.

Для описания таких данных лучше подходит не среднее, а медиана. Напомним, что медиана — это значение, которое делит распределение пополам: половина значений больше медианы, половина — меньше (точнее, не больше). Из рис. 2.6А видно, что ровно половина юпитериан выше 36 см. Стало быть, 36 см — это медиана роста юпитериан.

Для характеристики разброса роста юпитериан найдем значения, не выше которых оказались 25 и 75% результатов измерения.

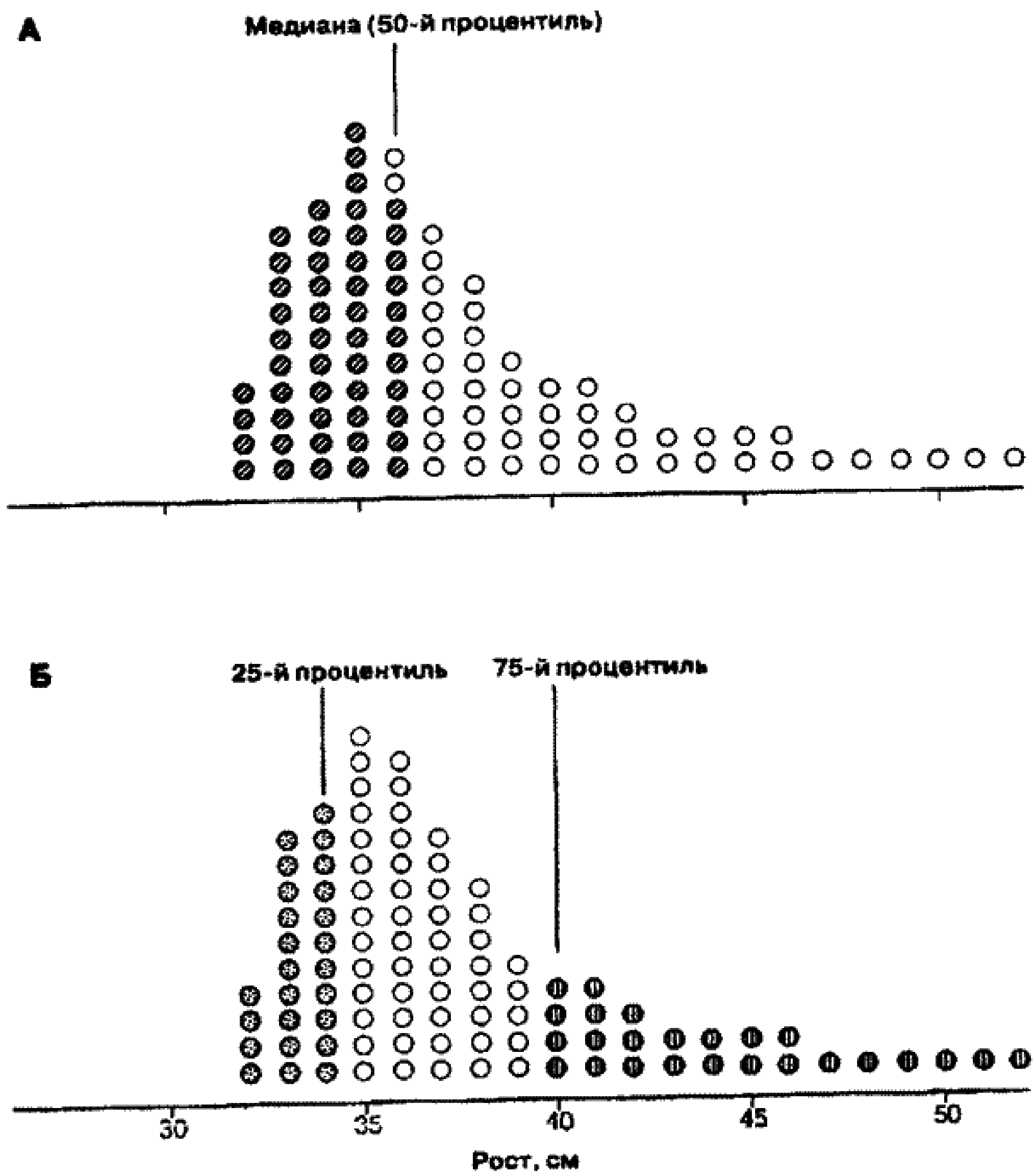


Рис. 2.6. Для описания асимметричного распределения следует использовать медиану и процентили. Медиана — это значение, которое делит распределение пополам. А. Медиана роста юпитериан —36 см. Б. 25-й и 75-й процентили отсекают четверть самых низких и четверть самых высоких юпитериан. 25-й процентиль ближе к медиане, чем 75-й — это говорит об асимметричности распределения

Эти величины называются 25-м и 75-м процентилями. Если медиана делит распределение пополам, то 25-й и 75-й процентили отсекают от него по четвертушке. (Саму медиану, кстати, можно считать 50-м процентилем.) Для юпитериан, как видно из рис. 2.6,Б 25-й и 75-й процентили равны соответственно 34 см и 40 см. Конечно, медиана и процентили, в отличие от среднего и стандартного отклонения, не дают полного описания распределения. Однако между 25-м и 75-м процентилями находится половина значений — значит, мы можем судить, каков ростом средний юпитерианин. По положению медианы относительно 25-го и 75-го процентилей можно судить о том, насколько асимметрично распределение. И наконец, теперь мы примерно знаем, кто на Юпитере считается высоким (выше 75-го процентиля), а кто ростом не вышел (ниже 25-го процентиля).

Для описания распределения чаще всего применяют 25-й и 75-й процентили. Однако можно рассчитывать любые другие процентили. Часто используют 5-й и 95-й процентили.

Вычисление процентилей — хороший способ разобраться в том, насколько распределение близко к нормальному. Напомним, что для нормального распределения 95% значений заключено в пределах двух стандартных отклонений от среднего и 68% — в пределах одного стандартного отклонения; медиана совпадает со средним (см. рис 2.4). Соответствие между процентилями и числом стандартных отклонений от среднего таково (см. также рис. 2.7).

Если соответствие между процентилями и отклонениями от среднего не слишком отличается от приведенного, то распределение близко к нормальному и его можно описать при помощи среднего и стандартного отклонения.

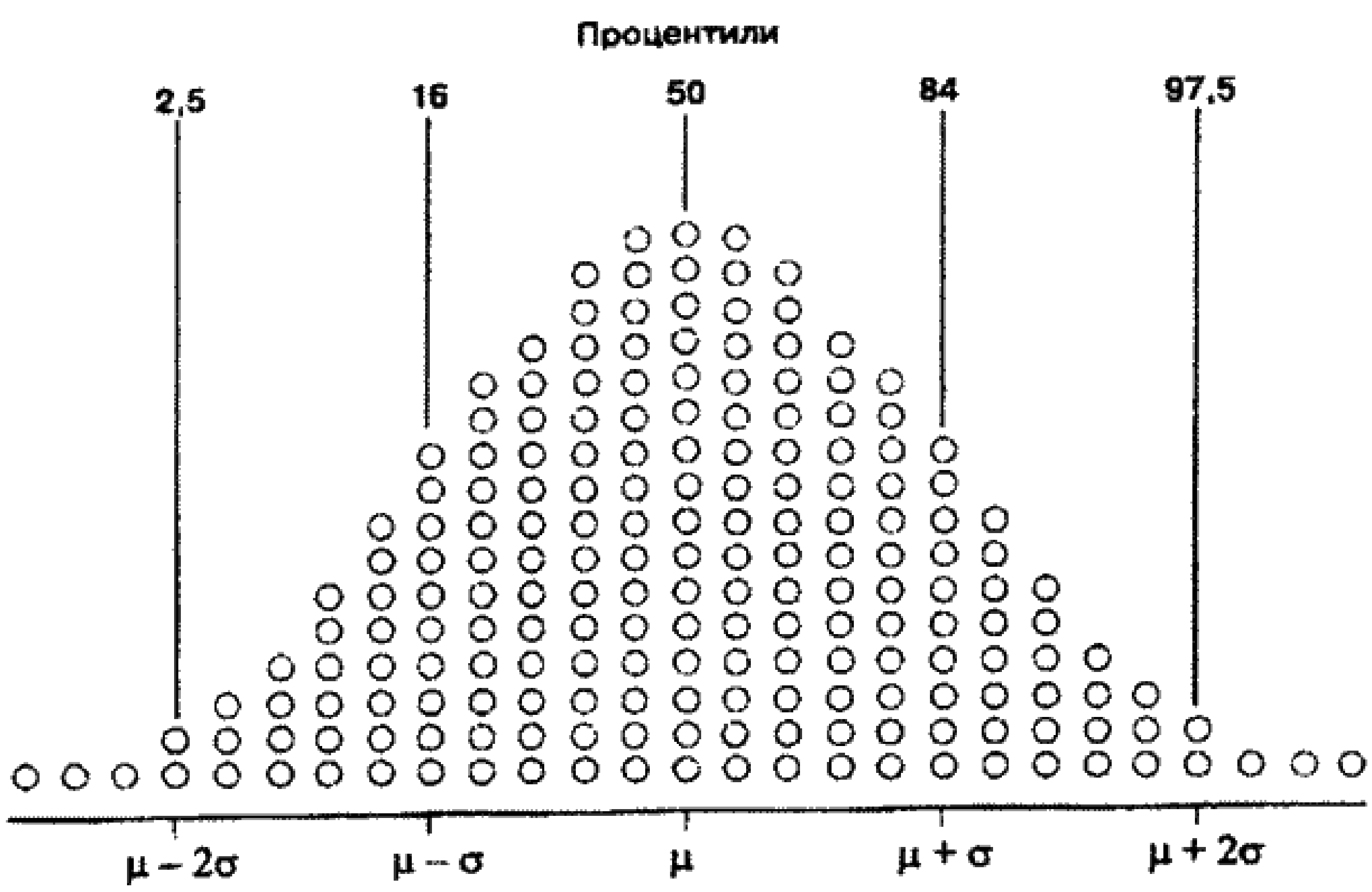


Рис. 2.7. Нормальное распределение: соответствие между числом стандартных отклонений от среднего и процентилями

Есть еще одна, и очень важная, причина, по которой нужно знать, близко ли распределение к нормальному. Дело в том, что многие методы проверки гипотез основаны на предположении, что распределение близко к нормальному. Только в этом случае эти методы будут надежны.

Здесь необходимо подчеркнуть, если известно, что выборка скорее всего принадлежит к совокупности с нормальным распределением, лучше всего использовать выборочное среднее и выборочное стандартное отклонение. Если есть основания полагать, что распределение в совокупности отличается от нормального, следует использовать медиану, 25-й и 75-й процентили.

Выборочное среднее и выборочное стандартное отклонение есть оценки среднего и стандартного отклонения для совокупности, вычисленные по случайной выборке. Понятно, что разные выборки дадут разные оценки. Для характеристики точности выборочных оценок используют стандартную ошибку.

Стандартную ошибку можно подсчитать для любого показателя, но сейчас мы остановимся на стандартной ошибке среднего — она позволяет оценить точность, с которой выборочное среднее характеризует значение среднего по всей совокупности.

На рис. 2.8.А представлено уже знакомое нам распределение марсиан по росту. Мы уже знаем рост каждого марсианина. Посмотрим, что получится, если оценивать средний рост по выборке объемом, скажем, 10 марсиан.

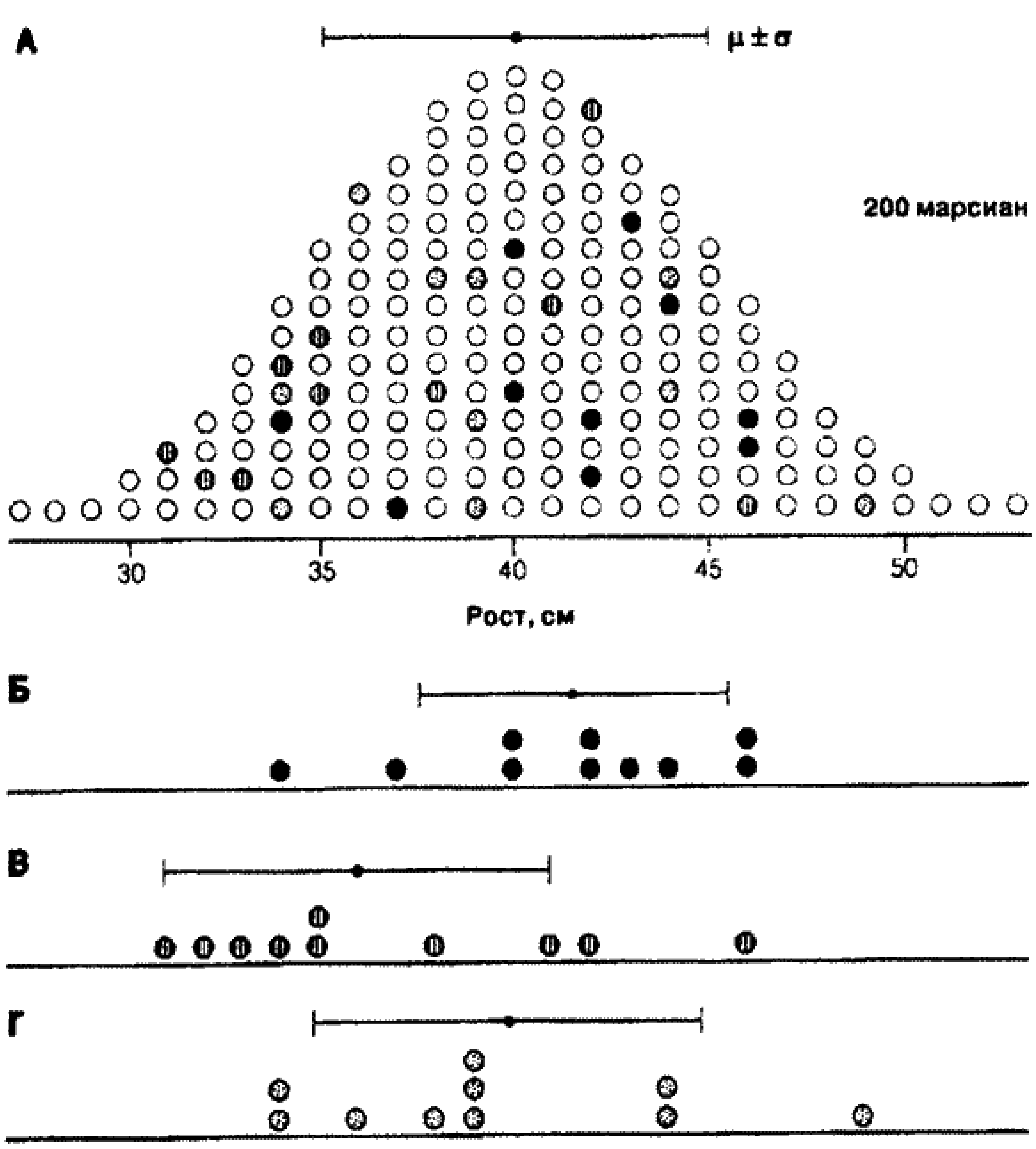


Рис. 2.8. Три случайные выборки из одной совокупности дают три разных оценки среднего и стандартного отклонения

Из 200 обитателей Марса наугад выберем 10 , пометим их черными кружками (рис. 2.8,А) и вынесим их на отдельный рисунок (рис. 2.8,Б). Точка и два отрезка по бокам от нее изображают выборочное среднее (X = = 41,5 см) и выборочное стандартное отклонение (s = 3,8 см). Эти значения близки, но не равны среднему по совокупности (μ = 40 см) и стандартному отклонению (σ = 5 см).

Извлечем еще одну случайную выборку того же объема. Результат показан на рис. 2.8,В. На рис. 2.8,А попавшие в эту выборку марсиане изображены заштрихованными кружками. Выборочное среднее (36 см) по-прежнему близко к среднему по совокупности, хотя и отличается от него; что касается выборочного стандартного отклонения (5 см), то на этот раз оно совпало со стандартным отклонением по совокупности.

На рис. 2.8,Г представлена третья выборка. Попавшие в нее марсиане на рис. 2.8,А изображены кружками с точками. Среднее и стандартное отклонение для этой выборки составляют соответственно 40 и 5 см.

Теперь пора поставить добычу случайных выборок на промышленную основу. Рассмотрим совокупность средних для каждой из возможных выборок по 10 марсиан. Общее число таких выборок превышает 10^{16} . Три из них мы уже обследовали. Средние по этим выборкам представлены на рис. 2.9 в виде заполненных кружков. Пустые кружки — это средние еще для 22 выборок. Итак, теперь каждому выборочному среднему соответствует кружок ,

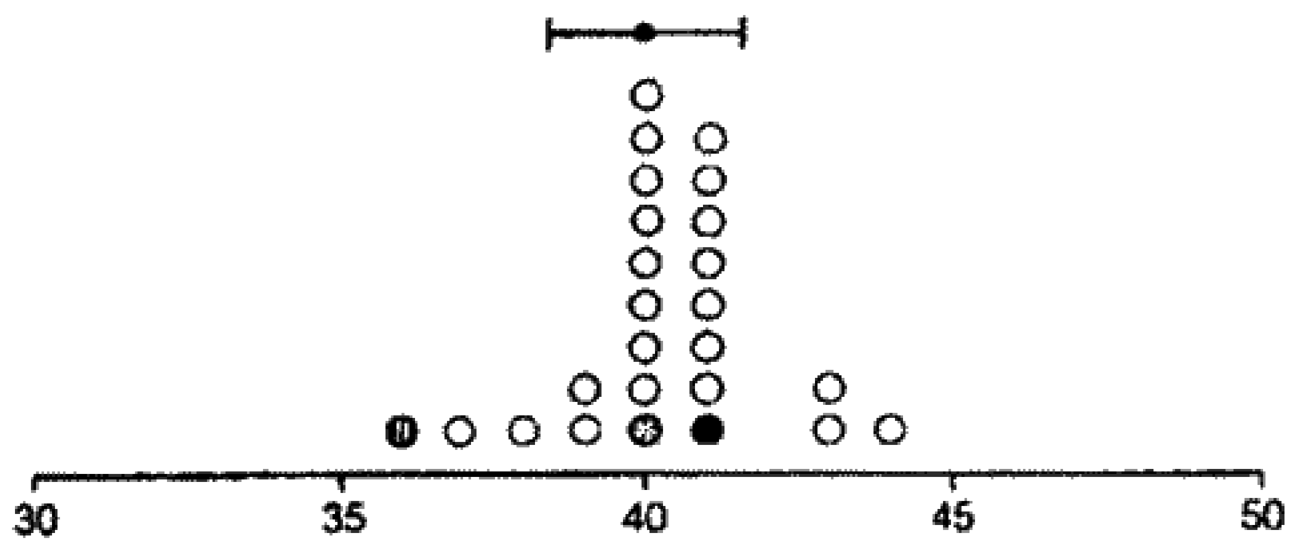


Рис. 2.9. Такое распределение мы получим, выбрав 25 раз по 10 марсиан из совокупности, представленной на рис. 2.8,А и рассчитав среднее для каждой выборки (средние для трех выборок с рис. 2.8 показаны заполненными кружками). Если построить распределение средних для всех возможных выборок, оно окажется нормальным. Среднее этого распределения будет равно среднему той совокупности, из которой извлекаются выборки. Стандартное отклонение этого распределения называется стандартной ошибкой среднего

точно так же, как до сих пор кружки соответствовали отдельному объекту.

Посмотрим на рис. 2.9. Набор из 25 выборочных средних имеет колоколообразное распределение, похожее на нормальное. Это не случайно. Можно доказать, что если переменная представляет собой сумму большого числа независимых переменных, то ее распределение стремится к нормальному, какими бы ни были распределения переменных, образующих сумму. Так как выборочное среднее определяется именно такой суммой, его распределение стремится к нормальному, причем чем больше объем выборок, тем точнее приближение. (Если выборки принадлежат совокупности с нормальным распределением, распределение выборочных средних будет нормальным независимо от объема выборок.)

Поскольку распределение на рис. 2.9 нормальное, его можно описать с помощью среднего и стандартного отклонения.

Так как среднее значение для рассматриваемых 25 точек есть среднее величин, которые сами являются средними значениями, обозначим его $\overline{X}_{\overline{X}}$.

Аналогично, стандартное отклонение обозначим $s_{\overline{X}}$. По формулам для среднего и стандартного отклонения находим: $\overline{X}_{\overline{X}}$=40 см и $s_{\overline{X}}$ = 1,6 см.

Среднее выборочных средних $\overline{X}_{\overline{X}}$ оказалось равно среднему μ всей совокупности из 200 марсиан. Ничего неожиданного в этом нет. Действительно, если бы мы провели исследования всех возможных выборок, то каждый из 200 марсиан был бы выбран равное число раз. Итак, среднее выборочных средних совпадет со средним по совокупности.

Интересно, равно ли $s_{\overline{X}}$ стандартному отклонению σ совокупности из 200 марсиан? Стандартное отклонение для совокупности выборочных средних $s_{\overline{X}}$ равно 1,6 см, а стандартное отклонение самой совокупности — 5 см. Почему $s_{\overline{X}}$ меньше, чем σ? В общих чертах это можно понять, если учесть, что в случайную выборку редко будут попадать одни только коротышки и одни гиганты. Чаще их будет примерно поровну, и отклонения роста от среднего будут сглаживаться. Даже в выборке, куда попадут 10 самых высоких марсиан, средний рост составит только 50 см, тогда как рост самого высокого марсианина —53 см. Подобно тому как стандартное отклонение исходной выборки из 10 марсиан s служит оценкой изменчивости роста марсиан, $s_{\bar{x}}$ является оценкой изменчивости значений средних для выборок по 10 марсиан в каждой. Таким образом, величина $s_{\bar{x}}$ служит мерой точности, с которой выборочное среднее $\bar{X}$ является оценкой среднего по совокупности μ. Поэтому $s_{\bar{x}}$ носит название стандартной ошибки среднего.

Чем больше выборка, тем точнее оценка среднего и тем меньше его стандартная ошибка. Чем больше изменчивость исходной совокупности, тем больше изменчивость выборочных средних; поэтому стандартная ошибка среднего возрастает с увеличением стандартного отклонения совокупности.

Истинная стандартная ошибка среднего по выборкам объемом n, извлеченным из совокупности, имеющей стандартное отклонение σ, равна:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}. \qquad (2.7)$$

Собственно стандартная ошибка — это наилучшая оценка величины $\sigma_{\bar{x}}$ по одной выборке:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = s\sqrt{1/n}, \qquad (2.8)$$

где s — выборочное стандартное отклонение.(Формулы 2.7, 2.8 даны без выводов).

Так как возможные значения выборочного среднего стремятся к нормальному распределению, истинное среднее по совокупности примерно в 95% случаев лежит в пределах 2 стандартных ошибок выборочного среднего.

Как уже говорилось, распределение выборочных средних приближенно всегда следует нормальному распределению независимо от распределения совокупности, из которой извлечены выборки. В этом и состоит суть утверждения, называемого центральной предельной теоремой:

- Выборочные средние имеют приближенно нормальное распределение независимо от распределения исходной совокупности, из которой были извлечены выборки.
- Среднее значение всех возможных выборочных средних равно среднему исходной совокупности.
- Стандартное отклонение всех возможных средних по выборкам данного объема, называемое стандартной ошибкой среднего, зависит как от стандартного отклонения совокупности, так и от объема выборки.

По мере того как мы увеличиваем объем выборки, выборочное среднее $\bar{X}$ и стандартное отклонение s дают все более точные оценки среднего μ и стандартного отклонения σ по совокупности. Увеличение точности оценки среднего отражается в уменьшении стандартной ошибки среднего $\sigma_{\bar{x}}$. Набрав достаточное количество марсиан, можно сделать стандартную ошибку среднего сколь угодно малой. В отличие от стандартного отклонения стандартная ошибка среднего ничего не говорит о разбросе данных — она лишь показывает точность выборочной оценки среднего.

Хотя разница между стандартным отклонением и стандартной ошибкой среднего совершенно очевидна, их часто путают (значение стандартной ошибки среднего заведомо меньше стандартного отклонения и исследователям кажется, что в таком виде их данные внушают больше доверия). Может быть, так оно и есть, однако беда в том, что стандартная ошибка среднего измеряет именно точность оценки среднего, но никак не разброс данных.

В заключение этого раздела мы хотели бы обратить внимание на следующее.

Если Вы откроете программу MS EXCEL для анализа данных и найдете раздел “описательная статистика”, то вы сможите просчитать знакомые уже Вам статистики , а также ряд других показателей ,характеризующих так называемый “вариационный ряд”. Этот ряд обычно состоит из конечного числа замеренных значений какого - либо признака (рост,вес и т.п) .

Помимо уже знакомых нам среднего, моды и т. д., к описательной статистике вариационного ряда относят такие характеристики, как асимметрия, эксцесс, интервал, мини – максимальные значения ряда, его сумму, количество значений в ряду и др.

Кратко рассмотрим некоторые из них.

Асимметрия (A_x)– коэффициент, характеризующий асимметричность распределения ряда относительно среднего.Изменяется от -1 до 1.Если асимметррия больше 0 ,то среднее значение находится правее моды (рис 2.1) ,если меньше 0, то среднее–левее моды, при асимметрии равной 0 распределение симметрично.

Эксцесс (E_x) - коэффициент , характеризующий “крутость “ распределения. При нормальном распределении E_x=0 , при E_x>0 – островершинное распределение, при $E_x < 0$ – пологое .

Здесь необходимо отметить , что $\overline{X}$, S^2, A_x, E_x еще называют ценральными моментами 1,2,3,4 порядка соответственно /2,7/.

Интервал (R) - разность между максимальным и минимальным значениями ряда. Используется редко (при очень маленьких выборках).

Более подробную информацию о вышеперечисленных характеристиках и формулы для расчетов можно найти в работах /2,4,7/.

Примеры расчета характеристик вариационного ряда приведены в четвертой части данного пособия.

Часть 3. ГИПОТЕЗЫ, КРИТЕРИИ, ЗНАЧИМОСТИ

Вернемся к нормальному распределению и рассмотрим в отличие от" галактического примера " диаметрально противоположный пример , связанный с наукой о Земле – палеонтологией, изучающей по остаткам организмов историю прошлых геологических эпох /4/. Рассмотрим выборку очень большой коллекции брахиопод рода Composita, с замеренными у них длиной и шириной раковин. (Брахиоподы - обитатели морского дна ,тела которых покрыты раковинами. Ведут прикрепленый образ жизни, систематизируются по видам и родам).

3.1 , Стандартизация

Предположим , что частотная диаграмма длин раковин будет выглядеть аналогично графику, изображенному на рис. 3.1. Среднему значению длины, в данном случае равному 14,2 мм, будет соответствовать наибольшая частота, а постепенно уменьшающимся и увеличивающимся значениям будут отвечать уменьшающиеся частоты. Приблизительно две трети раковин попадают в пределы интервала $(\mu - s, \mu + s)$ с центром в точке $\mu = 14{,}2$, причем оценка стандартного отклонения приблизительно равна 4,7 мм.

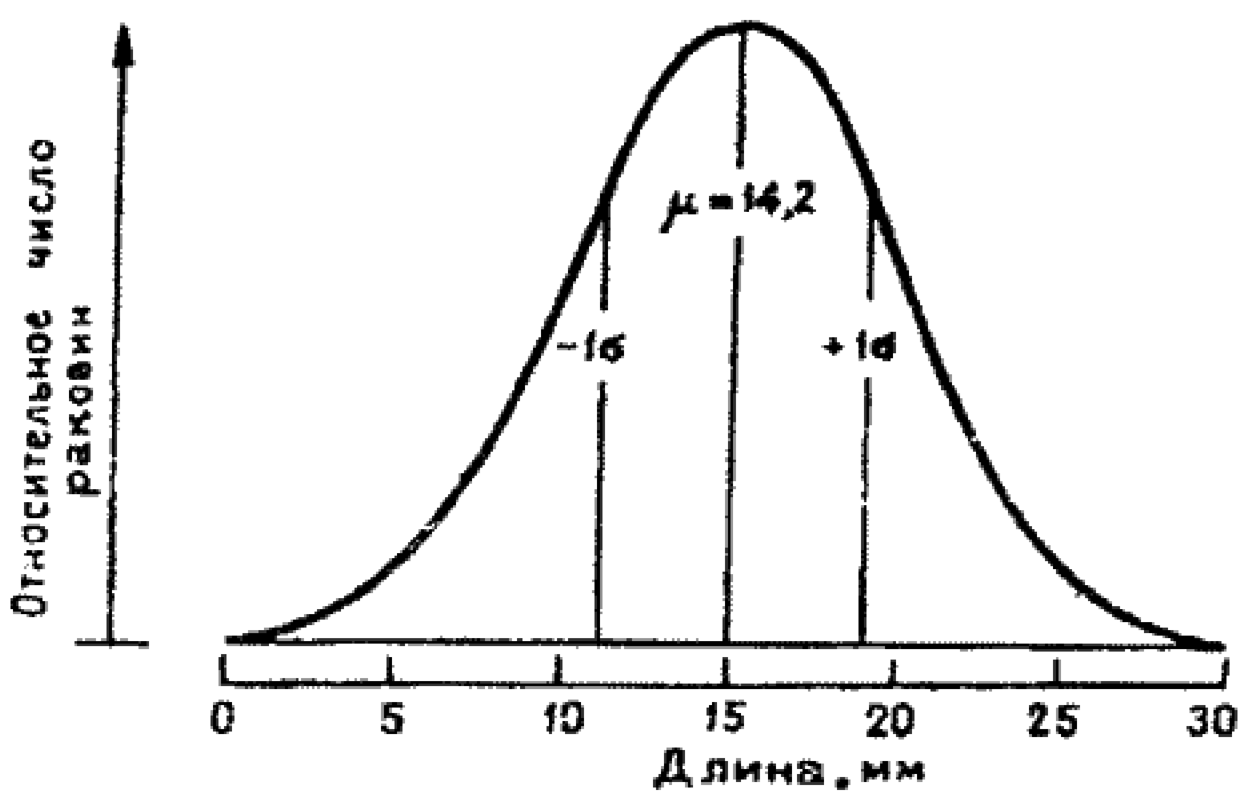


Рис. 3.1. Гипотетическое распределение значений длины особей рода Composita

Теперь рассмотрим измерения ширины, которые были сделаны при исследовании этой очень большой коллекции Composita. Распределение этого показателя по форме напоминает распределение длины, но его среднее значение и стандартное отклонение в этом случае иные. Оно может выглядеть, например, подобно графику, изображенному на рис. 3.2, со средним значением 10,3 мм и стандартным отклонением 3,6 мм.

Можем ли мы сравнивать два распределения друг с другом? Измерения проведены в одних и тех же единицах, что облегчает проблему сравнения распределений длины и ширины. Оба эти распределения можно изобразить в одном и том же масштабе, в результате чего получим график (рис. 3.3.)

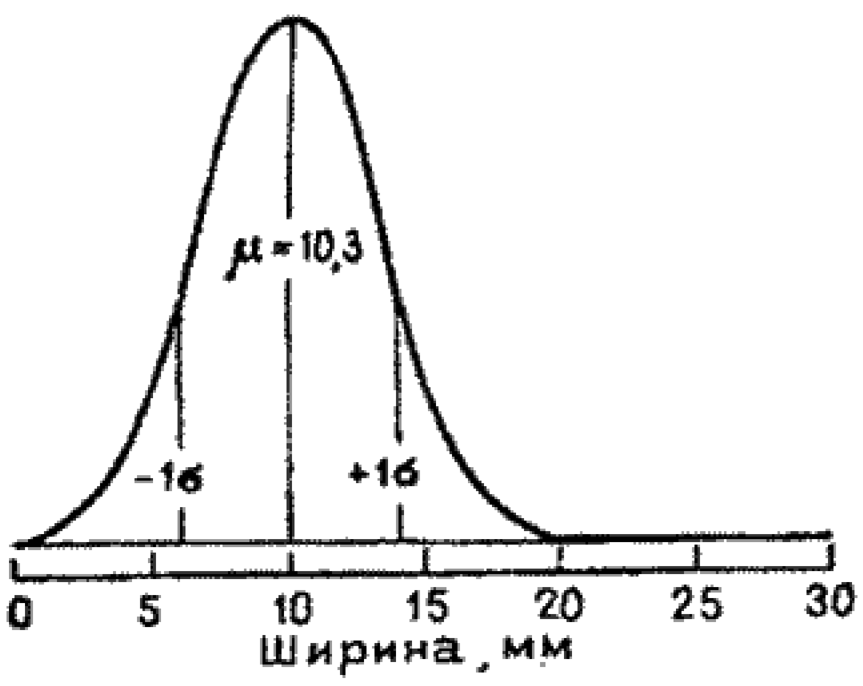


Рис. 3.2. Гипотетическое распределение значений ширины особей рода Composita

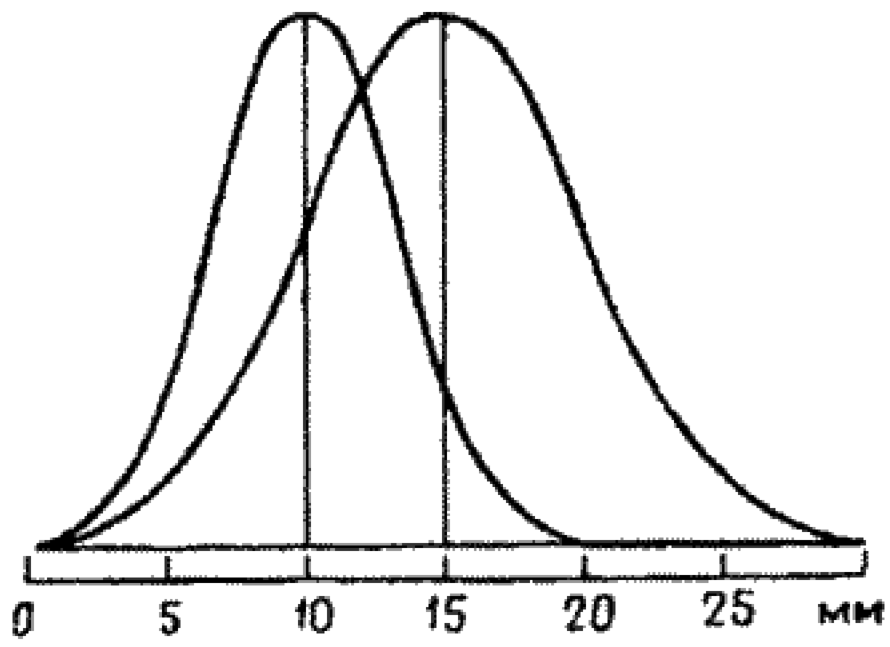


Рис. 3.3. Диаграмма распределения значений длины и ширины особей родаComposita

Конечно, сравнение было бы проще, если бы оба распределения имели один и тот же центр, т. е. равные средние значения. Мы можем центрировать их по отношению к общему среднему значению, вычитая подходящее число из всех значений совокупности (или прибавляя некоторое число к значениям другой совокупности) таким образом, чтобы средние обеих совокупностей совпали. Вместо этого вычтем соответствующее среднее значение из каждого наблюдения в каждой из двух совокупностей. Получим новые значения. Это преобразование сдвигает каждое из распределений вдоль горизонтальной оси до тех пор, пока их центры не совпадут со значением 0, являющимся средним для обоих преобразованных распределений (рис. 3.4) .

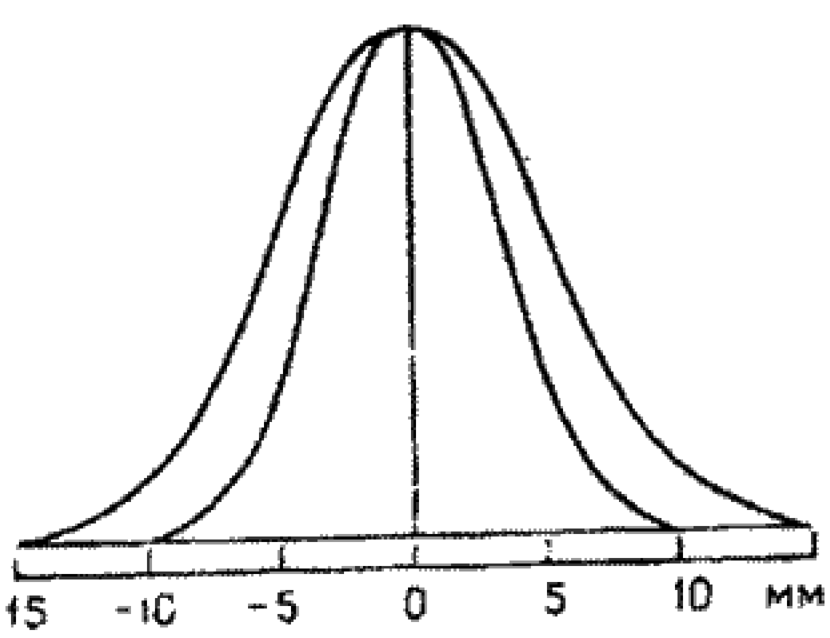


Рис. 3.4. Распределения значений длины и ширины особей рода Composita

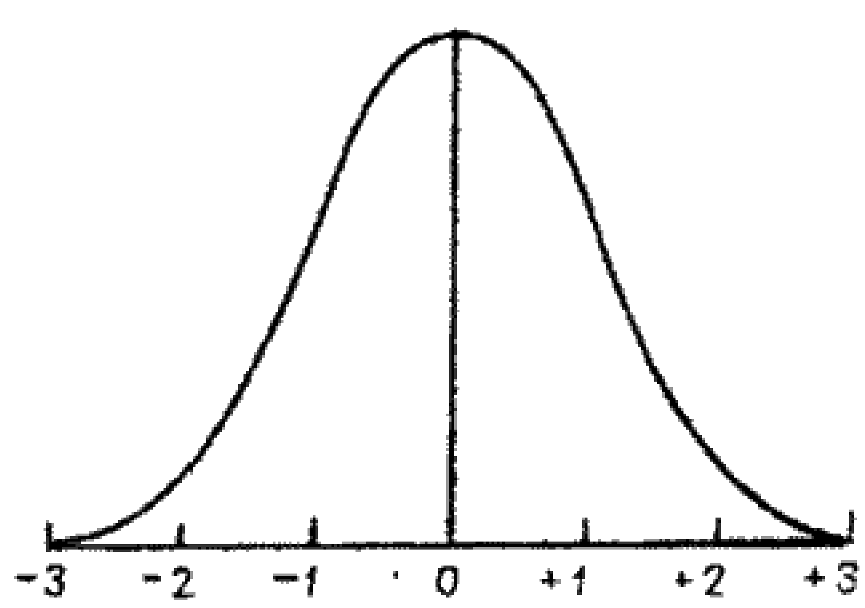


Рис 3.5. Распределения значений длины и ширины особей рода Composita после стандартизации имеют нулевое среднее и стандартное отклонение ,равное 1

В рассмотренном примере мы связаны размерностью результатов измерений, выраженной в миллиметрах. При этом никаких проблем не возникает, если мы будем сравнивать распределения длины и ширины, но если мы захотим сравнивать эти распределения с распределениями, характеризующими массу раковин, то нам это сделать не удастся. Существует ли какое-либо дополнительное преобразование, которое позволяет сделать наши распределения не зависящими от единиц измерения? Одно из таких чрезвычайно полезных преобразований называется стандартизацией: в результате его применения новые значения переменных имеют не только нулевое среднее значение, но также измеряются в единицах стандартных отклонений. Это делается просто с помощью вычитания среднего значения распределения из каждого наблюдения и деления каждой полученной разности на стандартное отклонение распределения. Эта новая переменная имеет стандартную нормальную форму и считается по формуле:

$$Z_i = \frac{(X_i - \overline{X})}{s}, \tag{3.1}$$

где $\overline{X}$, s - среднее значение и стандартное отклонение; X_i - результаты измерений распределения с исходной размерностью ; Z_i - стандартизованные значения исходного распределения со средним значением , равным нулю и стандартным отклонением равным единице (рис 3.5).

3.2. Z – критерий

Теперь (рис. 3.5) наши кривые частот различных совокупностей рода Composita идентичны. Характеристики стандартного нормального распределения очень хорошо известны, а таблицы площадей, ограниченных указанными сегментами кривой, можно найти почти во всех учебниках по статистике. Напомним, что площади выражаются прямо через вероятности. Используя таблицу (табл. 3.1), можно найти любую вероятность, связанную со случайной выборкой из нормальной совокупности, значения которой расположены в некотором заданном интервале. Однако для этого нужно знать дисперсию совокупности.

Продолжим работу с примером Девиса по брахиоподам и сделаем нереальное предположение, что мы исследовали всю совокупность рода Composita. Это значит, что мы знаем среднее значение длин ее элементов, равное 14,2 мм, и их стандартное отклонение, равное 4.7 мм. Какова вероятность появления при случайном выборе образца, меньшего 3 мм? Для получения ответа на этот вопрос стандартизуем 3 мм по формуле (3.1) и обратимся к табл. 3.1:

$$Z = \frac{(3.0 - 14.2)}{4.7} = -2.4$$

Таблица 3.1

Значения кумулятивной функции распределения стандартного нормального распредения /4/.

Стандартные отклонения от среднего значения	Кумулятивная вероятность	Стандартные отклонения от среднего значения	Кумулятивная вероятность	Стандартные отклонения от среднего значения	Кумулятивная вероятность
—3,0	0,0014	—0,9	0,1841	+1,1	0,8643
—2,9	0,0019	—0,8	0,2119	+1,2	0,8849
—2,8	0,0026	—0,7	0,2420	+1,3	0,9032
—2,7	0,0035	—0,6	0,2743	+1,4	0,9192
—2,6	0,0047	—0,5	0,3085	+1,5	0,9332
—2,5	0,0062	—0,4	0,3446	+1,6	0,9452
—2,4	0,0082	—0,3	0,3821	+1,7	0,9554
—2,3	0,0107	—0,2	0,4207	+1,8	0,9641
—2,2	0,0139	—0,1	0,4602	+1,9	0,9713
—2,1	0,0179	—0,0	0,5000	+2,0	0,9773
—2,0	0,0228	+0,0	0,5000	+2,1	0,9821
—1,9	0,0287	+0,1	0,5398	+2,2	0,9861
—1,8	0,0359	+0,2	0,5793	+2,3	0,9893
—1,7	0,0446	+0,3	0,6179	+2,4	0,9918
—1,6	0,0548	+0,4	0,6554	+2,5	0,9938
—1,5	0,0668	+0,5	0,6915	+2,6	0,9953
—1,4	0,0808	+0,6	0,7257	+2,7	0,9965
—1,3	0,0968	+0,7	0,7580	+2,8	0,9974
—1,2	0,1151	+0,8	0,7881	+2,9	0,9987
—1,1	0,1357	+0,9	0,8159	+3,0	0,9981
—1,0	0,1587	+1,0	0,8413		

Вероятность получения представителя совокупности рода Composita, длина которого меньше — 2,4 стандартных отклонений, есть кумулятивная вероятность в этой точке: по табл. 3.1 найдем значение 0,0082, которое в действительности очень мало. Теперь вычислим вероятность появления представителя, длина которого превышает 20 мм. Требуемую величину преобразуем в стандартную нормальную форму:

$$Z = \frac{(20.0 - 14.2)}{4.7} = 1.2$$

Так как суммарная площадь под кривой нормального распределения равна 1,00, то вероятность получения величины x, равной или большей 1,2 стандартных отклонений, т. е. большей, чем среднее, равна разности 1,00 и кумулятивной вероятности получения значений, не превосходящих 1,2. Иначе говоря,

$$P(x > 1{,}2) = 1{,}0 - P(x < 1{,}2).$$

Табл. 3.1 дает нам кумулятивные вероятности вплоть до 1,2, и вычитаемая вероятность равна 0,8849. Поэтому вероятность появления особей Composita длиннее 20 мм равна 1,0000— 0.8849= 0.1151, или немногим больше одной десятой. Теперь вычислим вероятность случайного выбора Composita, длина которой попадает в интервал от 15 до 20 мм:

для 15 мм

$$Z = \frac{(15.0 - 14.2)}{4.7} \approx 0.2$$

для 20 мм

$$Z = \frac{(20.0 - 14.2)}{4.7} \approx 1.2 \quad ,$$

$$P\ (x \le 1.2) = 0{,}8849,$$

$$P(x \le 0.2)\ = 0{,}5793,$$

$$P\ (0.2 \le x \le 1.2) = 0{,}3056,$$

т. е. примерно одна треть образцов попадает в заданный интервал.

(В качестве примера попробуйте определить вероятности выбора для различных обследований марсиан ростом 30 и 46 см и венерианцев от 10 до 15 см).

Вернемся к центральной предельной теореме, изложенной во втором разделе и где сказано ,что, к сожалению, мы обычно не знаем, какой вид имеет распределение, но иногда подозреваем, что оно значительно отличается от нормального. Из этого не следует, что нормальное распределение бесполезно, так как имеет место центральная предельная теорема. Она утверждает, что если выборки извлечены случайно из любой совокупности, то средние, вычисленные для этих данных, а именно выборочные средние, являются случайными величинами, распределение которых стремится к нормальному при увеличении объема выборки.

Центральная предельная теорема позволяет сформулировать статистические критерии, основанные на характеристиках нормальной кривой, и применять их даже в тех случаях, когда совокупность, из которой взята выборка, не распределена нормально. Предположим, что палеонтолог, который занимался исследованием коллекции Composita, нашел очень большую плиту, покрытую брахиоподами. Ископаемые выглядят аналогично Composita, но по размерам очень велики, средняя длина шести образцов составляет примерно 30,0 мм. Напомним, что мы "знаем", что среднее и стандартное отклонения совокупности Composita соответственно равны примерно 14,2 и 4,7 мм. Можно ли считать, что новая выборка брахиопод была извлечена из этой совокупности? Мы можем определить разность между средним значением нашей новой выборки и средним значением совокупности. Эту разность затем можно сравнить с изменчивостью, которую мы бы хотели иметь для средних значений выборок, случайно извлеченных с заданной совокупности. Эта изменчивость задается стандартной ошибкой и является функцией как дисперсии совокупности, так и объема выборки. (Уравнение 2.8).

Сравнение между разностью средних и стандартной ошибкой можно осуществить по следующей формуле

$$Z = \frac{\overline{X} - \mu}{s_{\bar{x}}} = \frac{\overline{X} - \mu}{s\sqrt{\frac{1}{n}}} \quad , \tag{3.2}$$

Заметим, что проверяемая статистика вычисляется таким образом, что она в точности эквивалентна критерию, используемому для стандартизации переменной (см. уравнение 3.1). Проверяемая статистика Z нормально распределена со средним значением, равным нулю, и стандартным отклонением, равным единице, если выборочное среднее действительно было получено для гипотетической совокупности. Если Z крайне велико, то мы вправе заключить, что наша выборка не была взята из этой совокупности. Формальное решение, однако, требует, чтобы мы установили соответствующую процедуру для вычисления проверяемой статистики.

3.3. Гипотезы

Первый шаг в статистической проверке гипотез — формулировка подходящей гипотезы об исследуемой переменной. Обычно такая гипотеза называется нулевой, обозначается H_0 и, в сущности, является гипотезой об отсутствии различия. Например, можно предположить, что данная выборка взята из совокупности, имеющей заданное среднее значение. Нулевая гипотеза выражается в форме

$$H_0 : \mu_1 = \mu_0, \tag{3.3}$$

которая означает, что среднее значение μ_1 изучаемой совокупности, из которой была взята выборка, равно заданному среднему значению.

В примере мы должны будем предположить, что среднее значение совокупности, из которой были взяты брахиоподы, находящиеся на плите, совпадает со средним значением совокупности рода Composita.
Сформулировав нулевую гипотезу, мы должны указать и альтернативу к ней. Подходящая альтернатива в этой ситуации может быть следующей:

$$H_1 : \mu_1 \neq \mu_0, \tag{3.4}$$

т. е. что среднее значение совокупности, из которой была взята выборка, не равно заданному значению μ_0.

Теперь рассмотрим процедуры проверки гипотез при заданном уровне значимости. Если две изучаемые совокупности окажутся различными, следует сделать вывод, что ископаемые остатки были взяты не из совокупности рода Composita, а из совокупности некоторого другого рода.

Как только гипотеза сформулирована, можно на основании нашего статистического критерия принять ее или отвергнуть. Гипотеза также может быть истинной или ложной. Это приводит к тому, что возникает четыре комбинации возможных исходов, две из которых приводят к правильному выводу, а две — к неправильному. Это можно проиллюстрировать следующим образом:

	Гипотеза верна	Гипотеза неверна
Гипотеза принимается	Правильное решение ,1- α	Ошибка второго рода , β
Гипотеза отвергается	Ошибка первого рода , α	Правильное решение, 1- β

Только принятие правильной или отклонение неправильной гипотезы можно считать верным решением. Если нулевая гипотеза отвергается , а на самом деле она верна, то возникает ошибка, называемая ошибкой первого рода. Наоборот, если ошибочная гипотеза принимается, то совершается ошибка второго рода.

Возвращаясь к нашему примеру, проиллюстрируем введенные понятия:

Гипотеза	В действительности	
	Особи с плиты принадлежат совокупности	Особи с плиты не принадлежат совокупности
μ плиты$=\mu_0$	Правильное решение	Ошибка второго рода
μ плиты$\neq\mu_0$	Ошибка первого рода	Правильное решение

Здесь " μ плиты " относится, конечно, к среднему значению совокупности, к которой принадлежат особи, собранные с плиты.

3.4. Уровни значимости , вероятности

В статистических процедурах вероятность появления ошибки первого рода обозначается α и называется уровнем значимости; эту вероятность можно задать до применения критерия. Для того чтобы минимизировать вероятность появления ошибки второго рода, запишем нулевую гипотезу при условии, что она будет отклонена. Если гипотеза отклоняется, то вероятность появления ошибки второго рода равна нулю, тогда как вероятность появления ошибки первого рода известна, так как она задается заранее. Если критерий не приводит к отклонению нулевой гипотезы (т. е. нулевая гипотеза принимается), то появляется некоторая вероятность сделать ошибку второго рода. Эта вероятность β, вообще говоря, неизвестна. Таким образом, если гипотеза о равенстве средних отвергается, мы делаем вывод о том, что две изучаемые совокупности имеют различные средние значения и вероятность того, что принято ошибочное решение, равна α. С другой стороны, если H_0 не отвергается, утверждение о том, что средние двух совокупностей совпадают, может оказаться ложным с неизвестной вероятностью β.

Все статистические критерии основаны на предположении, что нулевая гипотеза и альтернатива к ней взаимно исключают друг друга и вместе образуют полное множество событий. Так как нулевая гипотеза записывается в явном виде, то альтернатива должна быть довольно общей. Если H_0 отвергается, то мы считаем, что заданное соотношение, описываемое нулевой гипотезой, не выполняется. Более того, истинное соотношение в этом случае содержится в обширном множестве альтернатив, заключенных в общей альтернативе. Мы не можем определить какое из соотношений истинно; мы можем только установить, какое из соотношений не выполняется. Иногда в статистике применение статистических критериев позволяет говорить об "опровержении нулевой гипотезы" против альтернативы о неуспехе опровержения. Неуспех опровержения, которому соответствует неизвестная вероятность принятия ошибочного решения, не служит эквивалентом принятия гипотезы. Статистические критерии в некотором смысле не могут сказать нам, что именно имеет место, а только могут сказать, чего нет.

Возвращаясь к нулевой гипотезе и альтернативе, определенной формулами (3.3) и (3.4), предположим, что мы сочли уровень значимости (т. е. вероятность ошибки первого рода) α = 0,05 подходящим для наших целей. Иными словами, мы допускаем возможность приблизительно 5 раз на 100 испытаний ошибочно отвергнуть проверяемую гипотезу в случае, когда она верна.

Предположим, что дисперсия совокупности, по отношению к которой ведется проверка, нам известна. Палеонтолог определил, что дисперсия значений длины для совокупности особей рода Composita равна 22,1 (напомним, что стандартное отклонение было 4,7). Теперь можно формально записать статистический критерий следующим образом:

1) пусть проверяемая гипотеза и альтернатива имеют вид

$$H_0: \mu_1 = \mu_0,$$
$$H_1: \mu_1 \neq \mu_0;$$

2) принимаем уровень значимости: α=0,05;

3) вычисляем статистический критерий:

$$Z = \frac{\overline{X} - \mu_0}{\sigma\sqrt{\frac{1}{n}}} \quad (3.5)$$

Если выборка взята наудачу из нормальной совокупности с известной дисперсией, то статистический критерий Z будет распределен нормально со средним значением, равным нулю, и дисперсией, равной единице. Мы приняли соглашение о том, что приблизительно один раз на 20 (α=0.05) испытаний допускается ошибочное отклонение гипотезы о равенстве средних, в то время как она верна. Иными словами, мы принимаем 5 %-ный уровень риска или вероятность ошибки первого рода равную 0,05. Определим для стандартизованного нормального распределения область, заключающую 5% площади под кривой нормального распределения. Эта область называется критической. Если вычисленное значение статистического критерия попадает в эту область, мы вынуждены отклонить нулевую гипотезу.

Так как альтернатива—просто одно из неравенств, то гипотеза будет отклонена, если значение критерия слишком велико или слишком мало. Это значит, что существуют три возможные ситуации: $\mu_1 = \mu_0$; $\mu_1 > \mu_0$; $\mu_1 < \mu_0$.

В данном случае нас не интересует различие между двумя последними неравенствами. Критическая область охватывает крайние значения оси абсцисс, причем каждая подобласть занимает 2,5% площади, ограниченной кривой нормального распределения.

Сказанное можно резюмировать следующим образом: мы знаем характеристики нормальной кривой, которые получены из теоретических соображений, и поэтому их эмпирическое использование вполне оправданно. Если дисперсия нормально распределенной совокупности известна, то мы знаем также процентное содержание особей, размеры которых заключены в различных пределах (например, две трети особей приходится на интервал с центром в среднем значении, имеющий длину, равную двум стандартным отклонениям, рис .2.4). Если особи извлечены из этой совокупности случайным образом, вероятность получения выборки в заданном интервале кривой распределения равна площади, заключенной под соответствующей частью этой кривой. Если выборка взята из области, соответствующей очень малой вероятности, то это значит, что наша выборка не является выборкой из совокупности, указанной проверяемой гипотезой, которую мы отвергаем. Однако имеется некоторая вполне определенная вероятность извлечь выборку из критической области совокупности, равная площади этой критической области.

3.5 Оценивание. Доверительные интервалы

Возвращаясь к примеру рода Composita, напишем:

$$1)\ H_0 : \mu \text{ плиты } = 14.2 \text{ мм};$$
$$H_1 : \mu \text{ плиты } \neq 14.2 \text{ мм};$$
$$2)\ \alpha = 0.05;$$

$$3)\ Z = \frac{30.0 - 14.2}{4.7\sqrt{1/6}} \approx 8.2.$$

Мы уже знаем, что гипотеза о равенстве средних отвергается, если выборочное среднее либо слишком велико, либо слишком мало. Это приводит к двустороннему критерию на рис. 3.7. Критическая область, которая по соглашению должна содержать 5% площади нормального распределения, распадается

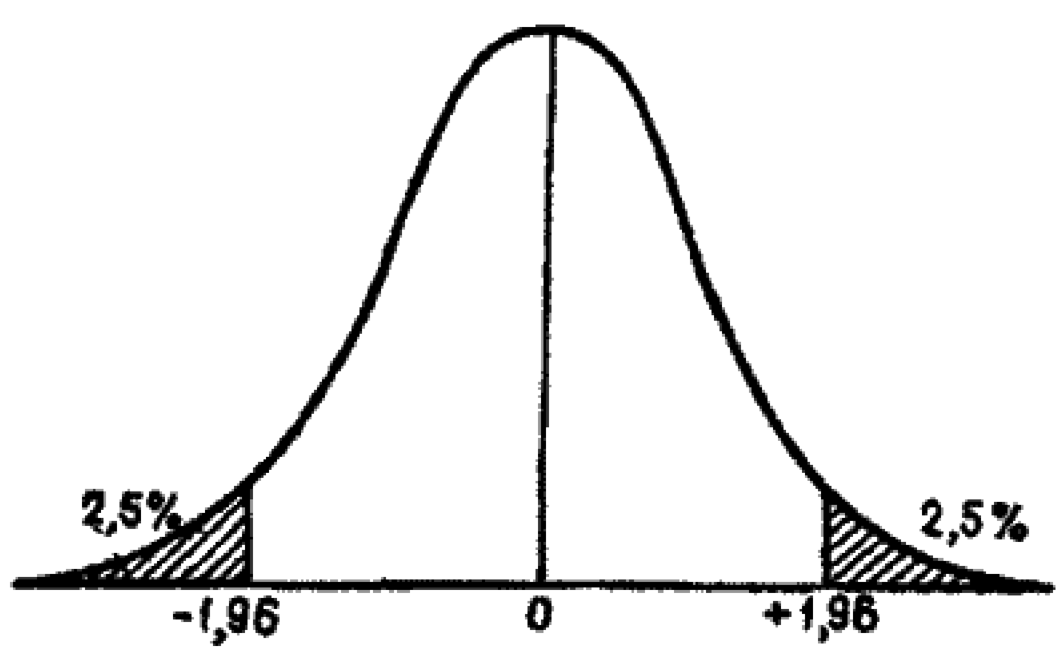


Рис. 3.7. Кривая нормального распределения с двумя заштрихованными критическими областями, охватывающими 5% площади под кривой

на две части, причем каждая из них содержит 2,5% общей площади. Если вычисленное значение Z попадает в левую половину, то мы делаем вывод, что выборка извлечена из совокупности, имеющей меньшее среднее значение, чем данная совокупность. Наоборот, если оно попадает в правую половину, то среднее выборочной совокупности больше, чем среднее заданной совокупности. Из табл. 3.1 мы находим, что приблизительно 2,5% площади под кривой находится слева от значения Z, равного - 1,9 (0,0287) и 97,5% (100%—2,5% =97,5%)— справа от значения +1,9 (0,9713).

Вычисленное значение критерия 8,2 превышает 1,9, из чего мы делаем вывод, что средние значения двух совокупностей не равны между собой, и коллекция ископаемых остатков на плите должна принадлежать к роду, отличному от рода Composita.

Необходимо отметить те допущения, которые делаются при использовании указанного критерия. Критерий Z основан на предположениях:

1) выборка брахиопод извлечена случайным образом;
2) совокупность длин остатков Composita распределена нормально;
3) дисперсия длин остатков Composita известна и равна 22,1 мм.

Если в частном примере какое-либо из указанных предположений является необоснованным, результаты, полученные с применением Z-критерия, могут показаться сомнительными. Тогда следует обратиться к другим критериям и процедурам принятия решений, основанных на предположениях, более отвечающих ситуациям .

Перед тем как перейти к дальнейшему изложению материала, связанному с анализом выборок различными способами, рассмотрению статистических критериев , гипотез и т. п., полезно сделать комментарий относительно выбора уровня значимости и привести примеры для более полного понимания вышеизложенного материала.

Часто (во многих руководствах по статистике) используются уровни значимости один к двадцати ($\alpha = 0{,}05$) или один к тысяче ($\alpha = 0{,}001$). Казалось, что подобная практика могла бы помочь обосновать целесообразность такого выбора, однако это не так. Определение уровня значимости находится целиком в компетенции исследователя, он должен решить, какой риск при отклонении истинной гипотезы является допустимым. Уровень значимости следует выбирать в соответствии с конкретными обстоятельствами, при которых используется критерий. Значения уровня α основываются на оценке последствий, которые возникнут, если сделать ошибку первого рода. Эти последствия могут быть очень важными и привести к потере денег, времени, здоровья , наконец, и даже жизней, или они могут быть неосязаемы и приводить к ущербу незначительному.

Обычно в статистических таблицах α предлагается от 0.001 до 0.9 или в процентах от 0.1% до 90%.

До сих пор неявно предполагалось, что совокупности и соответствующие им параметры известны (это и марсиане с венерианцами, и брахиоподы). Проблема, к которой мы попробуем перейти относится к ситуации, когда ничего не известно о совокупности, а имеющаяся информация относится только к самой выборке.

В действительности это гораздо более общая проблема. В социальных и в других опросах, в клинических испытаниях, в научных экспериментах полученные данные представляют собой выборки из много более обширных и неизвестных совокупностей. В конечном счете к выборкам и выводам из них относятся так же, как к информации о всей совокупности. Следовательно, следующим шагом должен быть анализ того, в какой мере выборочные результаты могут быть репрезентативны относительно совокупности.

Теперь, имея некоторые представления о совокупностях, выборках , их различных оценках, перейдем к примеру и рассмотрим выборку из 50 студентов, для которой рассчитан средний рост равный 174,94 см и стандартное отклонение - 6,42 см /9/. Если средняя совокупности неизвестна, то средняя выборки будет составлять единственную имеющуюся информацию относительно средней для совокупности. Поэтому выборочная средняя используется как оценка средней совокупности, так как другой информации нет. (Однако, хотя значение средней выборки может быть весьма близким к средней для совокупности, эти величины не обязательно совпадают).

Вариация в выборке и, следовательно, стандартное отклонение зависит от вариаций в совокупности. Если относительно совокупности неизвестно ничего, то такая информация также неизвестна. В подобных условиях вариация выборки может рассматриваться как мера вариации в совокупности, так как первая является производной от второй. Однако имеются некоторые сложности в использовании s - стандартного отклонения выборки как оценки σ, так как стандартное отклонение выборки смещено и неточно оценивает стандартное отклонение совокупности. Чем больше объем выборки, тем теснее зависимость между стандартными отклонениями выборки и совокупности. Если n последовательно растет, то выборочное стандартное отклонение приближается к стандартному отклонению совокупности. На практике эти стандартные отклонения достаточно близки, когда объем выборки по меньшей мере равен 30 /9/. В этом случае расхождением можно пренебречь. Поэтому в данном примере пойдет речь о выборках, где n - объем выборки больше или равен 30.

Стандартное отклонение s для среднего роста студентов в выборке из 50 человек известно — 6,42 см. Оно может служить оценкой стандартного отклонения совокупности σ и, следовательно, дает возможность оценить стандартное отклонение выборочного распределения $\sigma/\sqrt{n}$. Используя уравнение (2.8), имеем:

$$s/\sqrt{n} = 6.42 \text{ см.} / \sqrt{50} = 0.91 \text{ см.}$$

Таким образом, стандартное отклонение разброса возможных средних совокупности около средней выборки будет 0.91 см для выборки из 50 студентов со средним ростом 174.94 см.

Основываясь на свойствах нормального распределения, при заданных средней и стандартном отклонении выборки можно получить, что из возможных значений средней для совокупности приблизительно

68% находятся в интервале: выборочная средняя $\pm\, s/\sqrt{n}$;

95% находятся в интервале: выборочная средняя $\pm 1.96\, s/\sqrt{n}$;

99% находятся в интервале: выборочная средняя $\pm 2.58\, s/\sqrt{n}$;

99,9% находятся в интервале: выборочная средняя $\pm 3.3\, s/\sqrt{n}$.

Это представлено на рис. 3.8 (сравните рис. 3.8 с рис. 2.4).

Следовательно, средняя совокупности оценивается как выборочная, а приведенные значения являются индикаторами возможных колебаний от оценки. Так как приблизительно 68% возможных средних совокупности попадает в интервал $\overline{X} \pm s/\sqrt{}$, а 95% - в интервал $\overline{X} \pm 1.96\, s/\sqrt{n}$ и т.д., с вероятностью 0.68 исходная совокупность имеет среднюю в интервале $\overline{X} \pm s/\sqrt{n}$, с вероятностью 0,95 - в интервале $\overline{X} \pm 1.96\, s/\sqrt{n}$ и т. д.

Средняя для роста студентов в совокупности оценивается, следовательно, как 174,94 см. Кроме того, возможные вариации между истинной средней совокупности и ее оценкой 174,94 см уточняются следующим образом: из всех совокупностей, из которых могла бы быть извлечена данная выборка, приблизительно:

68% имеют средние в интервале 174,94 см ± 0,91 см, или от 174,03 см до 175,85 см;

95% имеют средние в интервале 174,94 см ± 1,78 см, или от 173,16 см до 176,72 см;

99% имеют средние в интервале 174,94 см ± 2,35 см, или от 172,59 см до 177, 29 см;
99,9% имеют средние в интервале 174,94 см ± 3,00 см, или от 171,94 см до 177, 94 см.

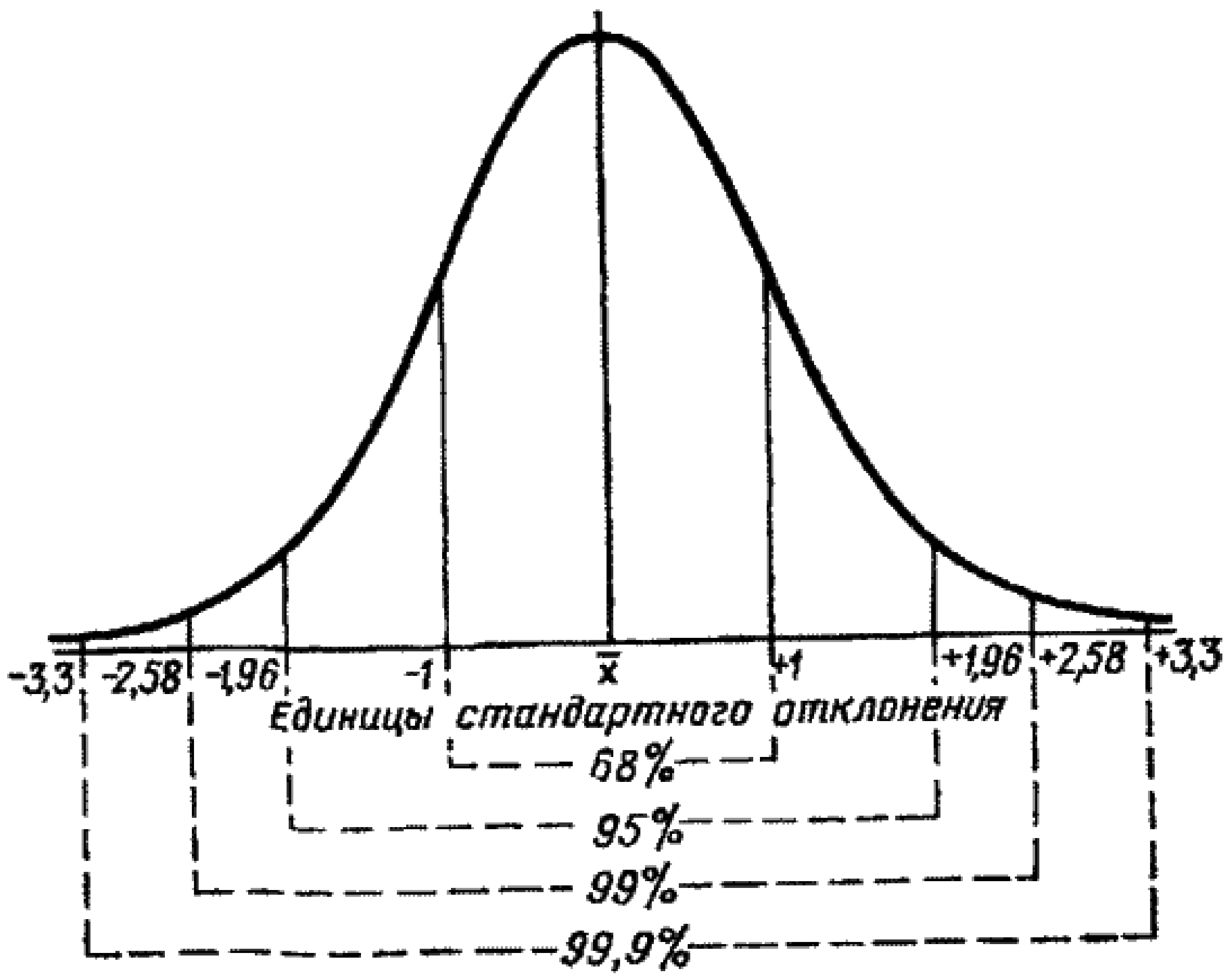


Рис. 3.8

Эти выводы можно выразить в вероятностных терминах: например, вероятность того, что средняя для роста студентов в совокупности лежит между 173,16 см и 176,72 см, равна 0.95, или вероятность того, что она находится между 172,59 см и 177,29 см, равна 0.99.

Эти выводы применимы только в тех случаях, когда *п* (объем выборки) достаточен (здесь он равен 50). К выборкам с $n \leq 30$ мы вернемся.

Не вдаваясь в рассуждения по определению доверительных границ (или пределов), покажем "что это такое" на примере со студентами 95% студентов имеют средние в интервале 174,94 см ± 1,78 см, или от 173,16 см до 176,72 см, то есть при оценке среднего роста студентов в совокупности в 174.94 см и по выборке, включающей 50 студентов, применяемый метод оценки обеспечивает 95% уровень доверия утверждению, что истинная средняя совокупности лежит между 173.16 см и176.72 см,

99 % - ный уровень предполагает доверительные границы в пределах от 172.59 см до 177.29 см.

Здесь мы можем сделать следующий вывод:
- увеличение вероятности (от 0.95 до 0.99) обусловливает расширение области, в которую должна попасть средняя совокупности (от 173.16 – 176.72 см до 172.59 – 177.29 см). Ценой увеличения надежности (вероятности) является потеря точности результатов (расширение границ для среднего совокупности) .

Теперь еще раз о доверительных границах (как о вероятной ошибке):
- доверительные границы указывают с данной вероятностью область тех значений, в которую, вероятно, попадет средняя для совокупности.

Другой путь формулировки того же результата состоит в указании вероятной ошибки при использовании выборочного результата в качестве оценки средней для совокупности. Тогда ошибка с вероятностью 0,95 есть величина, равная +1,96 стандартной ошибки, а ошибка с вероятностью 0,99 равна ±2,58 стандартной ошибки.

По выборке, состоящей из 50 студентов, установлено, что средний рост их в совокупности оценивается в 174,94 см. Ошибка с вероятностью 0,95 равна ±1,78 см (1,96*0,91 см), а ошибка с вероятностью 0,99 равна ±2,35 см (2,58*0,91 см).

Практически сложились специфические значения доверительных границ (главным образом 95%, 99% и 99,9%). Было найдено, что эти пределы "работают". т. е. они обеспечивают интерпретируемые результаты. 95%-ные доверительные границы удобны для тех данных, для которых трудно ожидать высокую степень точности. В эту категорию входит большинство социальных и экономических данных. Если потребуется большая степень определенности, то можно воспользоваться 99%-ными доверительными границами. В случае статистических данных более точной природы, где ожидаемые вариации слабы и где желательна большая определенность, предпочтительнее 99,9%-ные доверительные границы.

Мы уже знаем, что вопрос о том, принадлежит выборка данной совокупности или нет, может исследоваться при применении так называемой нулевой гипотезы (3.3). Это гипотетическое утверждение о том, что между выборочной статистикой (например, средней) и соответствующим параметром совокупности нет значимой разности. Следовательно, если имеется наблюденная разность, то предполагается, что она представляет собой результат случая, т. е. ошибки выборки. Цель состоит в том, чтобы обеспечить возможность для исследования нулевой гипотезы, иначе говоря, возможность принять или отвергнуть гипотезу. Если нулевая гипотеза отвергается, то наблюденная разность рассматривается как значимая или статистически значимая (чтобы подчеркнуть зависимость от применяемой статистической методологии).

Вернемся к примеру со студентами и рассмотрим выборку, включающую 50 студентов, средний рост которых равен 174,94 см, а стандартное отклонение — 6,42 см, и, предположим, что выборка извлечена из совокупности студентов, средний рост которых 172,50 см. Наблюденное значение разности между выборочной средней и средней совокупности равно: $\overline{X} - \mu = (174,94 - 172,50) =$ 2,44 см. Насколько вероятно, что эта разность случайна? Если уровень значимости равен $\alpha = 0,05$, это эквивалентно выяснению, попадет ли выборочная средняя в пределы 95%-ных доверительных границ или будет лежать вне этих пределов. Наиболее простой путь исследования этой проблемы состоит в преобразовании наблюденной разности в стандартные единицы (в значения Z) делением на соответствующую стандартную ошибку (3.2). Это значение Z можно сравнить с критическим значением, определяемым выбранным уровнем значимости. Здесь уровень значимости 0,05, так что критическое значение для сравнения с Z равно 1,96. Если уровень значимости 0,01, то критическое значение равно 2,58, а для уровня 0,001 оно равно 3,3 (рис3.8) .

Мы знаем, что для 50 студентов :

$$s/\sqrt{n} = 6.42/\sqrt{50} = 0.91 \text{ см.}$$

Соответствующее значение в стандартных единицах для наблюденной разности между средними определяется по (3.2) и

$$\frac{(174.94 - 172.50)}{0.91} = \frac{2.44}{0.91} = 2.68 \text{.}$$

При уровне значимости $\alpha = 0{,}05$ критическое значение Z для разности превосходит 1.96. Это означает, что нулевая гипотеза отвергается, так как вероятность того, что она верна, меньше, чем 0,05. Разность такой величины, как наблюденная (или больше), имеет вероятность менее чем 0,05 быть случайной, т. е. наблюдаться вследствие случайной ошибки. В действительности в этом примере значение Z превосходит 2.58 (рис.3.8) и, следовательно, выборочное значение случайно менее чем в 1 случае из 100 (уровень 0,01), а при $\alpha = 0.001$ выборочное значение не случайно,
так как $2.68 < 3.3$ (рис. 3.8)

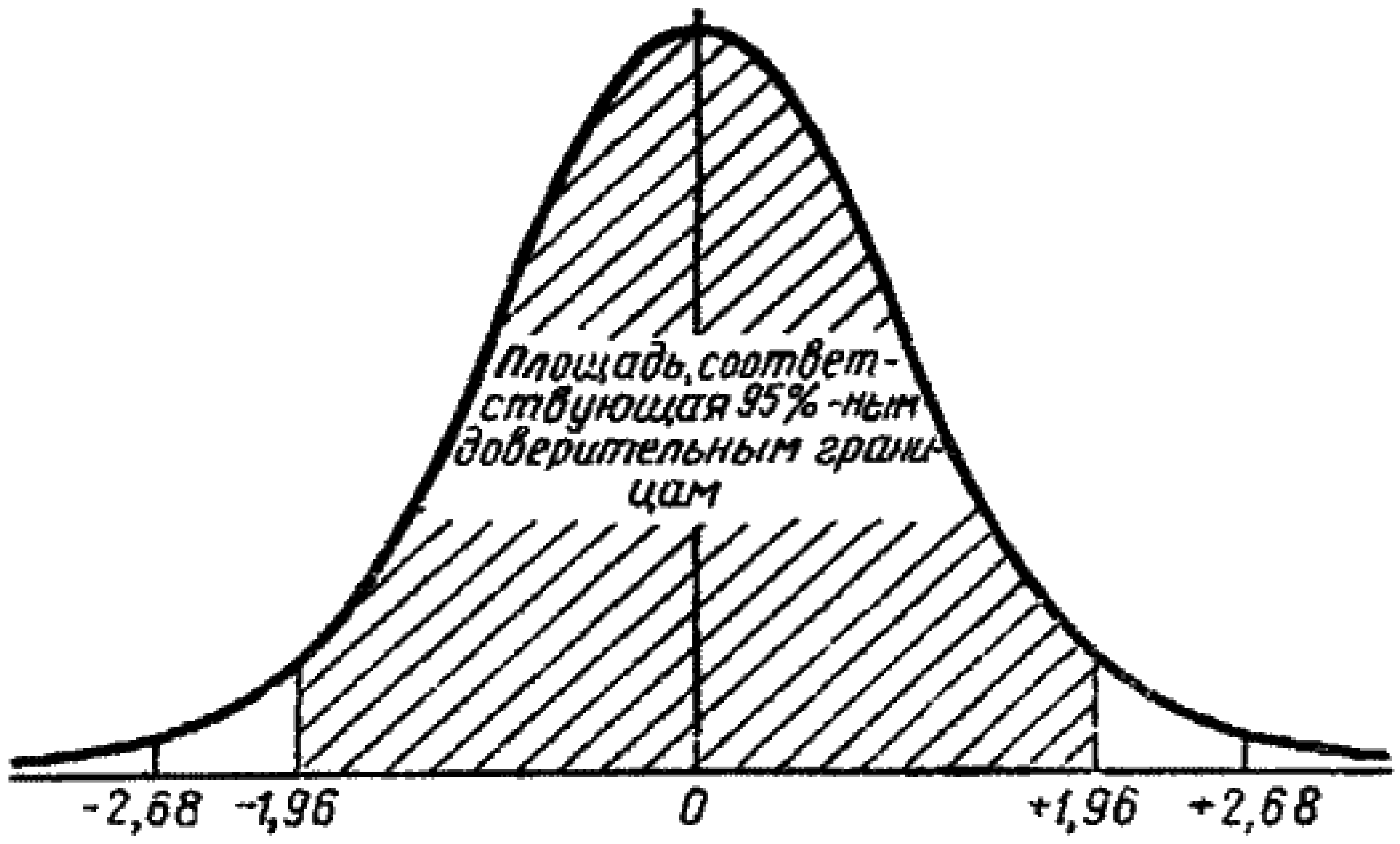


Рис 3.9. Соотношение между выборочной средней и средней совокупности в терминах выборочного распределения и критического значения Z

Полезно подчеркнуть, что вычисление значения Z для наблюденной разности устанавливает соотношение между двумя средними в терминах выборочного распределения. Это обеспечивает непосредственное знание вероятности, с которой может наблюдаться данное значение разности. В этом отношении указанный подход гораздо более предпочтителен, чем вычисление доверительных границ и выяснение положения выборочной средней относительно этих пределов.

Рассмотрим применение этого метода по этапам.

1. Предлагается нулевая гипотеза, состоящая в том, что разность между $\overline{X}$ и μ незначима и что любое наблюденное значение разности случайно или обусловлено случайной ошибкой.
2. Выбирается подходящее значение уровня значимости, например 0.05. Оно соответствует критическому значению $z = 1.96$.

Уровень значимости в некоторых изданиях по статистики, помимо известного обозначения ($\alpha = 0.05$ или $\alpha = 0.001$), обозначают как $z_{0.05}$ или $z_{0.001}$.

3. Устанавливается критерий, отвергающий нулевую гипотезу: она отвергается, если $|z| \geq z_{0.05}$ (1.96), где z вычисляется по формуле (3.2).
4. Вычисляется значение z. В этом примере для выборочной средней и средней для совокупности.

5. Нулевая гипотеза отвергается, если для вычисленного значения z $|z| \geq z_{0.05}$, и не отвергается, если $|z| < z_{0.05}$.

6. Если нулевая гипотеза отвергается, то вывод состоит в том, что выборка со средней $\overline{X}$ не принадлежит совокупности со средней μ. Разность между двумя средними значима.

Вернемся еще раз к примеру с 50 студентами, средний рост которых 174,94 см и "пройдемся" с ним по предложенным пунктам .

1. Выбранный уровень значимости равен $\alpha = 0.05$, так что критическое значение $z = 1.96$.

2. Нулевая гипотеза состоит в том, что разность между выборочной средней группы из 50 студентов (174,94 см) и предполагаемой средней для исходной совокупности (172,50 см) незначима. Наблюденная разность случайна, она обусловлена случайной ошибкой.

3. Нулевая гипотеза будет отвергнута, если

$$|z| \geq z_{0.05} \quad \text{или} \quad |z| \geq 1.96$$

4. Вычисленное значение z по формуле (3.2) равно 2.68 (вычисление см. выше)

5. $z = 2.68$, отсюда вытекает, что $|z| > z_{0.05}(1.96)$ и нулевая гипотеза отвергается.

6. Вывод, следовательно, состоит в том, что при уровне значимости $\alpha = 0.05$ выборка, включающая 50 студентов, средний рост которых 174,94 см, может не принадлежать совокупности студентов со средним ростом 172,50 см.

3.6. t – критерий. Степени свободы

Для того чтобы применить описанный выше Z – критерий, как мы уже знаем, нужно выполнить ряд условий, которые редко осуществимы на практике. Мы обычно не знаем истинных значений параметров изучаемого распределения, так как не можем изучить всей совокупности рода Composita, и ясно, что это нельзя сделать.

Так как μ и σ неизвестны, то лучшее, что можно сделать, — это оценить их по выборке, как это было проделано в примере со студентами . Однако такие оценки допускают некоторую степень неопределенности, поэтому решения, принимаемые на их основе, нельзя считать точными.

Неопределенность, возникающую как следствие применения оценок, построенных по выборке, можно учесть, если использовать распределение с более широкой областью значений, чем у нормального распределения. Одно из распределений такого типа называется t-распределением Стьюдента /4,6,7,8,9/.

Оно похоже на нормальное, но зависит от объема взятой выборки. Типичная кривая этого распределения изображена на рис. 3.10. Форма кривой меняется в зависимости от числа наблюдений. Когда число наблюдений в выборке бесконечно, то t - распределение совпадает с нормальным.

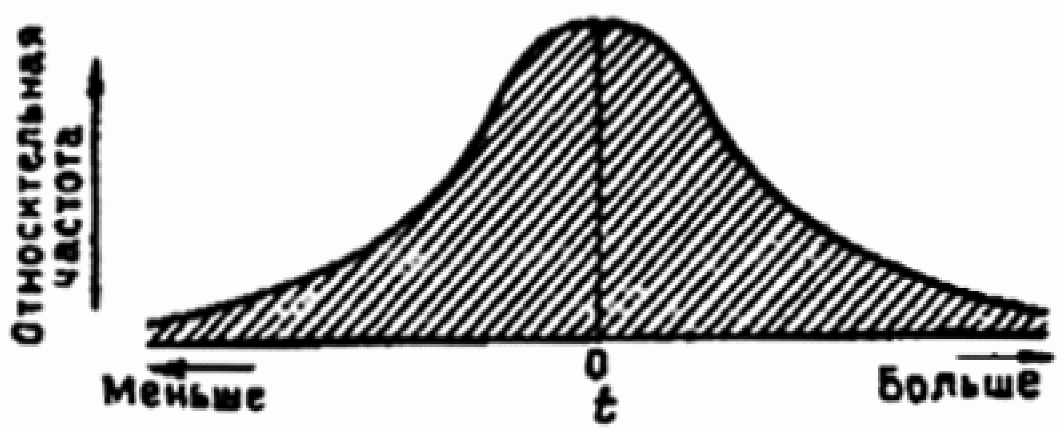


Рис.3.10 . t-распределение Стьюдента

Для того чтобы подсчитать значения статистического критерия, нужно по выборочным данным оценить параметры изучаемой совокупности. Интуитивно кажется невозможным решить сразу две задачи: оценить параметры совокупности и применить критерий, используя одну и ту же выборку без какой-либо компенсации, связанной с двукратным обращением к имеющемуся набору наблюдений. В связи с этим вводится величина, называемая числом степеней свободы, которую можно определить как разность между числом наблюдений в выборке и числом параметров, которые требуется оценить по выборочным данным. Иными словами, число степеней свободы — превышение числа наблюдений над числом оцениваемых параметров распределения. Числа степеней свободы обозначаются греческой буквой ν, это всегда целые положительные числа.

В качестве примера рассмотрим рис. 3.11, на котором представлено вычисление среднего и дисперсии выборки.

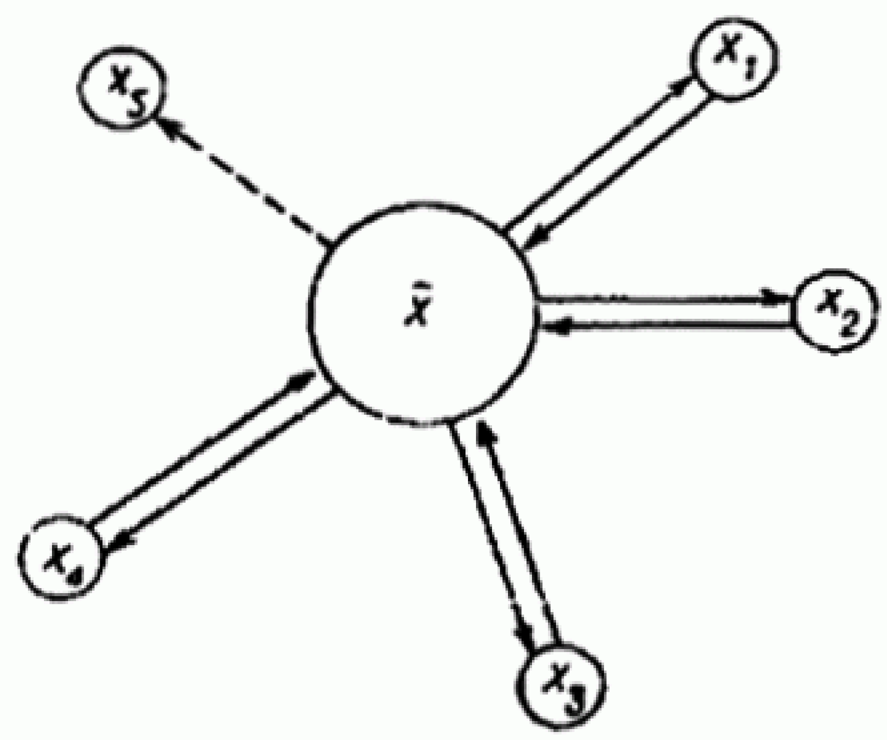


Рис. 3.11. Представление в виде диаграммы метода вычисления среднего и дисперсии по пяти наблюдениям. Среднее $\overline{X}$ вычисляется по всем наблюдениям; дисперсия – по разности между наблюдениями и средним. Когда четыре разности найдены, пятая разность известна

Среднее оценивается по пяти независимым наблюдениям и поэтому имеется пять степеней свободы. Дисперсия оценивается по пяти квадратам разностей $(\overline{X} - x_i)^2$. Однако заметим, что если мы определили четыре из этих разностей, то автоматически можно вычислить пятую, так как

$$\overline{X} - x_5 = 5\overline{X} - (x_1 + x_2 + x_3 + x_4).$$

Поэтому имеется только четыре зависимых источника информации, по которым вычисляется дисперсия.

К сожалению, понятие степеней свободы редко объясняется в начальных курсах статистики, скорее оно представляется как очевидное произвольное число, например $n-1$. Мы вкратце будем дальше рассматривать причины различия чисел степеней свободы, ассоциированных с различными статистическими критериями по мере их появления в тексте.

Таблицы t-распределения (и других выборочных распределений) используются точно таким же образом, как и таблицы кумулятивного стандартного нормального распределения; отличие состоит лишь в том, что для нахождения требуемой вероятности в таблице t- распределения надо знать два числа: α - заданный уровень значимости (вероятность ошибки первого рода) и число степеней свободы ν. Табл. 3.2 является таблицей t - распределения; более подробные таблицы можно найти во многих руководствах по математической статистике.

Так называемые t - критерии, которые основаны на распределении Стьюдента, полезны для проверки гипотезы о том, что данная выборка взята из совокупности с заданными характеристиками или же для проверки гипотезы об однородности двух выборок.

Проблемы такого типа являются основными в экспериментальных науках .

Пусть, например, нужно проверить гипотезу, заключающуюся в том, что ряд особей брахиопод, находящихся на одной из плит морского дна бассейна 1, результаты взвешивания которых приведены в табл. 3.3, взят из совокупности, имеющей средний вес более 18 грамм. Допустив, что особи были взяты наудачу из нормальной совокупности, вычислим t - критерий:

$$t = \frac{\overline{X} - \mu_0}{s_x} = \frac{\overline{X} - \mu_0}{s\sqrt{1/n}}, \tag{3.6}$$

где $\overline{X}$ - среднее арифметическое, вычисленное по данным выборки; μ_0 - гипотетическое среднее, равное 18 гр.; n - число наблюдений;

Критические значения t – критерия (двусторонний вариант) при ν степенях свободы и заданном уровне значимости /6/.

s - оценка стандартного отклонения; s_x - стандартная ошибка определения среднего значения. Заметим, что этот критерий совпадает с критерием (3.5), исключая то, что нужно оценить стандартную ошибку по формуле $s_x = s\sqrt{1/n}$, а не по формуле $\sigma\sqrt{1/n}$, так как истинная дисперсия совокупности неизвестна.

Таблица 3.2

Критические значения t – критерия (двусторонний вариант) при ν степенях свободы и заданном уровне значимости /6/.

	Уровень значимости, α								
ν	0,50	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	1,000	3,078	6,314	12,706	31,821	63,657	127,321	318,309	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
31	0,682	1,309	1,696	2,040	2,453	2,744	3,022	3,375	3,633
32	0,682	1,309	1,694	2,037	2,449	2,738	3,015	3,365	3,622
33	0,682	1,308	1,692	2,035	2,445	2,733	3,008	3,356	3,611
34	0,682	1,307	1,691	2,032	2,441	2,728	3,002	3,348	3,601
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
36	0,681	1,306	1,688	2,028	2,434	2,719	2,990	3,333	3,582
37	0,681	1,305	1,687	2,026	2,431	2,715	2,985	3,326	3,574
38	0,681	1,304	1,686	2,024	2,429	2,712	2,980	3,319	3,566
39	0,681	1,304	1,685	2,023	2,426	2,708	2,976	3,313	3,558
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551

v	Уровень значимости α 0,50	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
42	0,680	1,302	1,682	2,018	2,418	2,698	2,963	3,296	3,538
44	0,680	1,301	1,680	2,015	2,414	2,692	2,956	3,286	3,526
46	0,680	1,300	1,679	2,013	2,410	2,687	2,949	3,277	3,515
48	0,680	1,299	1,677	2,011	2,407	2,682	2,943	3,269	3,505
50	0,679	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
52	0,679	1,298	1,675	2,007	2,400	2,674	2,932	3,255	3,488
54	0,679	1,297	1,674	2,005	2,397	2,670	2,927	3,248	3,480
56	0,679	1,297	1,673	2,003	2,395	2,667	2,923	3,242	3,473
58	0,679	1,296	1,672	2,002	2,392	2,663	2,918	3,237	3,466
60	0,679	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
62	0,678	1,295	1,670	1,999	2,388	2,657	2,911	3,227	3,454
64	0,678	1,295	1,669	1,998	2,386	2,655	2,908	3,223	3,449
66	0,678	1,295	1,668	1,997	2,384	2,652	2,904	3,218	3,444
68	0,678	1,294	1,668	1,995	2,382	2,650	2,902	3,214	3,439
70	0,678	1,294	1,667	1,994	2,381	2,648	2,899	3,211	3,435
72	0,678	1,293	1,666	1,993	2,379	2,646	2,896	3,207	3,431
74	0,678	1,293	1,666	1,993	2,378	2,644	2,894	3,204	3,427
76	0,678	1,293	1,665	1,992	2,376	2,642	2,891	3,201	3,423
78	0,678	1,292	1,665	1,991	2,375	2,640	2,889	3,198	3,420
80	0,678	1,292	1,664	1,990	2,374	2,639	2,887	3,195	3,416
90	0,677	1,291	1,662	1,987	2,368	2,632	2,878	3,183	3,402
100	0,677	1,290	1,660	1,984	2,364	2,626	2,871	3,174	3,390
120	0,677	1,289	1,658	1,980	2,358	2,617	2,860	3,160	3,373
140	0,676	1,288	1,656	1,977	2,353	2,611	2,852	3,149	3,361
160	0,676	1,287	1,654	1,975	2,350	2,607	2,846	3,142	3,352
180	0,676	1,286	1,653	1,973	2,347	2,603	2,842	3,136	3,345
200	0,676	1,286	1,653	1,972	2,345	2,601	2,839	3,131	3,340
∞	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905

Формально мы проверяем гипотезу

$$H_0 : \mu_1 \le \mu_0$$

при множестве альтернатив

$$H_1 : \mu_1 > \mu_0.$$

Таблица 3.3

Результаты замера веса особей брахиопод с плиты бассейна 1

Номер особи	Вес, г
01	13
02	17
03	15
04	23
05	27
06	29
07	18
08	27
09	20
10	24

Сумма 213

Среднее 21.3

$s^2 = 30.46; s = 5.52; s = 1.75$

Проверяемая гипотеза заключается в том, что среднее значение веса совокупности, из которой была взята выборка, меньше или равно заданному значению 18 грамм. Множество альтернатив заключается в том, что изучаемая совокупность имеет средний вес, превосходящий 18 грамм.

Для определения критического значения t по табл. 3.2 требуется задать два числа: уровень значимости и число степеней свободы. В данном примере предполагается, что один параметр (μ) известен, а другой требуется оценить (оценкой для σ является величина s, т. е. выборочное стандартное отклонение). Поэтому выборке, содержащей десять измерений веса, соответствуют девять степеней свободы.

Нулевая гипотеза отвергается только в том случае, когда средний вес существенно превышает 18 грамм, и поэтому попадающими в критическую область можно считать только очень большие значения критерия, как это показано на рис. 3.12. Такой критерий называется односторонним, так как его критическая область расположена только с одной стороны области

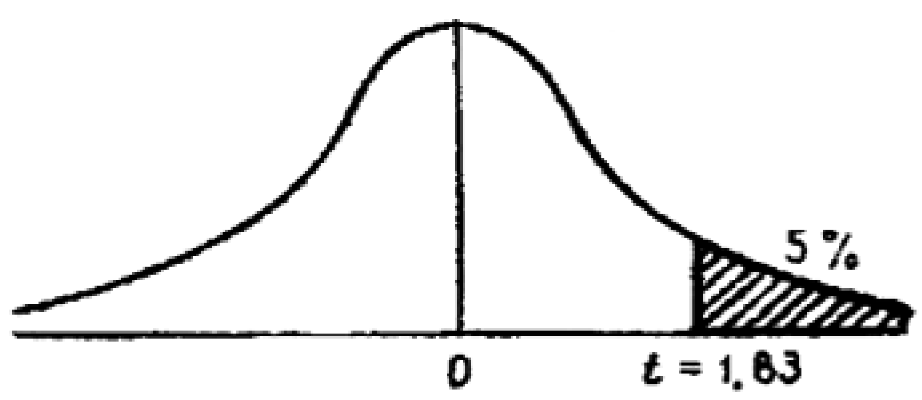


Рис. 3.13. Распределение Стьюдента с девятью степенями свободы

значений распределения. Если же нам нужно проверить эту гипотезу при уровне значимости α = 0.05, то вычисленное значение статистики t для одностороннего критерия должно превышать значение 1.83 (табл.3.2).

(В табл.3.2 для данного примера нужно смотреть $\alpha = 0.10$, так как вариант таблицы для двустороннего t- критерия, а пример для одностороннего критерия.)

Статистический критерий примет следующий вид:

1) $H_0 : \mu_1 \le 18$ г.

$H_1 : \mu_1 > 18$ г;

2) $\alpha = 0.05$;

3) $$t = \frac{21.3 - 18.0}{5.52\sqrt{1/10}} = 1.89.$$

Вычисленное значение 1.89 превышает табличное, соответствующее девяти степеням свободы и 5%-ному уровню значимости, т. е. попадает в критическую область. Это значит, что мы должны отклонить нулевую гипотезу и принять альтернативу, заключающуюся в том, что вес совокупности, из которой были взяты особи брахиопод, больше 18 г. Если бы вычисленная величина t оказалась меньше чем 1.83, то не было бы никаких оснований предполагать, что выборочное среднее больше 18 г. Заметим, что мы при этом не утверждаем, что среднее меньше 18 г. а только говорим, что нет оснований считать, что оно больше. Ранее было установлено, что эта неопределенность лежит в основе статистических критериев. Они могут показать с некоторой вероятностью, чего нет, но не позволяют установить, что же имеет место.

С плиты морского дна бассейна 2 были получены десять дополнительных измерений значений веса особей брахиопод в(табл. 3.4). Можно ли средние двух выборок считать равными?

В отличие от предыдущей задачи, где мы сравнивали выборочное среднее с заданным выборочным средним значением совокупности, в данном случае проверяется гипотеза, имеющая следующий вид:

$$H_0 : \mu_1 = \mu_2 .$$

Проверяемая гипотеза заключается в том, что среднее значение совокупности, из которой взята первая выборка, равно среднему значению совокупности, из которой взята вторая выборка. Множество альтернатив для гипотезы

$$H_1 : \mu_1 \neq \mu_2$$

утверждает, что средние значения двух совокупностей не равны. Снова мы должны задать уровень значимости, и пусть он будет равен 10% ($\alpha = 0.10$). Теперь статистический критерий имеет следующий вид:

$$t = (\overline{X_1} - \overline{X_2}) / s_e , \qquad (3.7)$$

где s_e – оценка стандартного отклонения разности между $\overline{X_1}$ и $\overline{X_2}$, полученная по двум объединенным выборкам.

Таблица 3.4

Результаты замера особей брахиопод с плиты бассейна 2.

Номер образца	Вес, г
11	15
12	10
13	15
14	23
15	18
16	26
17	24
18	18
19	19
20	21

Сумма 189

Среднее 18,9

$s^2 = 23.21$; $s = 4.81$;

Эту оценку s_e можно вычислить формуле

$$s_e = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} ,$$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}, \qquad (3.8)$$

где индексы соответствуют выборкам из бассейнов 1 и 2 . Процесс объединения двух выборок приводит к дополнительным степеням свободы, так как требуется оценить два параметра σ_1^2 и σ_2^2. Число степеней свободы поэтому для t –критерия эквивалентности, заданного формулой (3.6), есть $\nu = n_1 + n_1 - 2$. Является ли различие между двумя средними значимым при десятипроцентном уровне значимости?

$$s_p^2 = \frac{9(30.46)+9(23.21)}{10+10-2} = \frac{483.03}{18} = 26.84;$$

$$s_p = 5.18;$$

$$t = \frac{21.3-18.9}{5.18\sqrt{1/10+1/10}} = \frac{2.4}{2.32} = 1.03.$$

Так как табличные значения двустороннего критерия с 18 степенями свободы, соответствующие 10%-ному уровню значимости (5% на каждом конце распределения), равны —2.10 и 2.10 (табл 3.2), то вычисленное значение не попадает в критическую область и нулевую гипотезу нельзя отклонить. (Напомним, что критическая область охватывает 10% площади под кривой t -распределения.) Отсюда следует, что нет оснований предполагать, что две изучаемые выборки взяты из совокупностей, имеющих разные средние значения.

Для того чтобы применять этот критерий, необходимо выполнить следующие условия.

Во - первых, обе выборки должны быть получены на основании процедуры случайного выбора.

Во - вторых, значения случайных величин в совокупностях, из которых были извлечены выборки, должны описываться нормальным распределением.

В - третьих, дисперсии этих совокупностей должны быть равны.

Выполнение первого условия в большинстве задач проверить трудно. Однако его невыполнение в случае, если выборки имеют сильное и систематическое смещение (пример с проницаемостью в разделе 1), может явиться серьезным источником ошибок. Конечно, проверку гипотезы о нормальности распределения значений признака изучаемой совокупности можно провести, однако одно только отклонение от нормальности редко приводит к изменению результатов, в особенности если выборочная совокупность достаточно велика. Третье условие - равенство дисперсий двух совокупностей - очень важно, так как почти все статистические критерии основаны на предположении о равенстве дисперсий сравниваемых совокупностей. К счастью, это предположение легко проверяется.

Но, перед переходом к рассмотрению $F-$ критерия для проверки гипотезы о равенстве дисперсий , вернемся к выборкам с малым объемом .

3.7. Средняя для малых выборок

При рассмотрении примера о студентах, упоминалось об объеме выборки больше или равной 30. Теперь, имея представление о t – распределении, рассмотрим выборку с объемом $n \le 30$.

Значение объема выборки в 30 элементов как пограничное между малыми и большими выборками состоит в том, что при n=30 t-распределение очень тесно аппроксимируется ("описывается") нормальным и поэтому вариациями вследствие объема выборки можно пренебречь /9/.

Для объема выборки, меньшего чем 30, доверительные границы шире и вероятная ошибка больше, чем для выборок с объемом больше 30. При уменьшении объема выборки доверительные границы расширяются и вероятная ошибка возрастает.

В конце концов, для очень малых выборок доверительные границы столь широки, а вероятные ошибки столь велики, что практическая ценность любого статистического вывода незначительна.

Рассмотрим случайную выборку, состоящую из10 студентов. Средний рост студентов для нее равен 176.10 см, а стандартное отклонение равно 3.88 см. При 9 степенях свободы ($n-1$, где $n=10$) и вероятности 0.05 (это означает, что площадь каждого "хвоста" соответствующего t- распределения равна 0.025) $t=2.262$.

95%-ные доверительные границы для среднего роста студентов в совокупности даются соотношением:

$$\overline{X} \pm {ts}/\sqrt{n} \text{ или } 176.1 \pm 2.262 * 3.88/\sqrt{10} \text{ см,}$$

что равно 176.10 ± 2.78 см, т. е. от 173.32 до 178.88 см. Это больше соответствующих пределов (от 173.16 до 176.72см) для выборки, состоящей из 50 студентов (при $\alpha = 0.05$).

3.8. F -критерий

Критерии для проверки гипотезы о равенстве дисперсий основаны на так называемом F-распределении Фишера. Это теоретическое распределение отношения $F = s_1^2 / s_2^2$ двух выборочных дисперсий для выборок, взятых из нормальных совокупностей при условии, что истинные дисперсии равны.

Вполне естественно, что выборочные дисперсии в случае; когда число наблюдений, используемое для их вычисления, мало изменяются от испытания к испытанию в довольно широком диапазоне. Поэтому вид F - распределения изменяется с изменением объема выборки. Это снова заставляет учитывать степени свободы, но в данном случае F - распределение зависит от двух значений ν, каждое из которых соответствует одной из двух оценок дисперсий F – отношения. Так как F- статистика является отношением двух положительных чисел, то ясно, что случайная величина F не может принимать отрицательных значений. Если выборка велика, то при условии равенства истинных значений дисперсий среднее значение отношения будет близко к 1.0.

Нулевая гипотеза утверждает, что изучаемые совокупности имеют равные дисперсии; множество альтернатив устанавливает, что это не так. Степени свободы ν_1 и ν_2, отвечающие этому критерию, соответственно равны $n_1 - 1$ и $n_2 - 1$. Критическое значение F с ν_1=9 и ν_2=9 степенями свободы и уровнем значимости 5% (α=0.05) можно найти по табл. 3.5. Это значение равно 3.18. Значение F, вычисленное по формуле (3.9), попадает в одну из двух областей. Если вычисленное значение F превышает 3.18, то нулевая гипотеза отвергается и мы приходим к заключению, что дисперсии веса можно считать неодинаковыми в двух группах. Если вычисленное значение меньше 3.18, то мы не можем утверждать, что дисперсии различны. В качестве примера вычислим F – критерий веса двух совокупностей брахиопод и проверим предположение о равенстве дисперсий при 5 % -ном уровне значимости ($F = {}^{30.46}\!/_{28.21} = 1.07$). Вычисленное значение F не превышает табличного (3.18) и нулевая гипотеза принимается.

В большинстве практических задач мы обычно не знаем истинных значений параметров совокупности и можем лишь по выборке вычислить их оценки. При сравнении двух выборок сначала целесообразно установить, являются ли их дисперсии статистически эквивалентными и только убедившись в их эквивалентности можно переходить к использованию других статистических критериев (того же t – критерия).

Таблица 3.5

Критические значения F – распределения с ν_1 и ν_2 степенями свободы и 5% - ным уровнем значимости ($\alpha = 0.05$) /4/

ν_2	ν_1						
	1	2	3	4	5	6	7
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88
6	5,99	5,14	4,76	4,53	4;39	4,28	4,21
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01

Продолжение табл. 3.5

ν_1							
8	9	10	12	15	20	24	∞
238,88	240,54	241,88	243,91	245,95	248,01	249,05	250,10
19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46
8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62
6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75
4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50
4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81
3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38
3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08
3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86
3,07	3,02	2,98	2,91	2,84	2,77	2,74	2,70
2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57
2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47
2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,48
2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31
2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25
2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19
2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15
2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11
2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07
2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04
2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01
2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98
2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96
2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,04
2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92
2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90
2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88
2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87
2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85
2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84
2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74
2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65
2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55
1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46

Примеры расчетов статистических показателей, рассмотренных в третьей части приведены в четвертой части данного пособия.

4. Использование MS EXCEL в геолого –статистических расчетах

Вероятность

На рис. 4.1 показан пример определения вероятности в **MS EXCEL** по задаче первой части пособия об успешности бурения продуктивных скважин .
(При работе с Excel нужны первичные знания , которые можно получить из пособия /10/ или других работ).

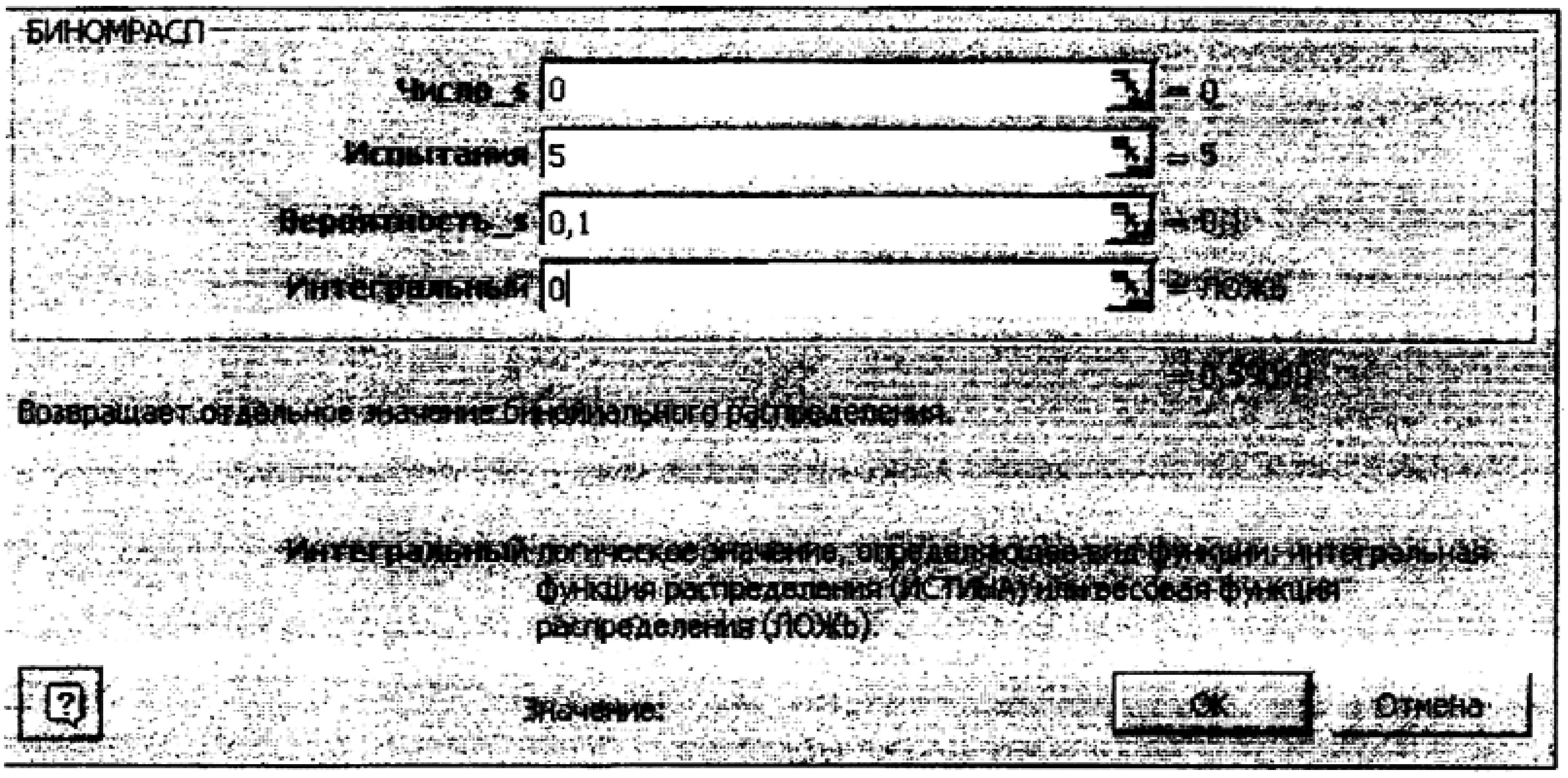


Рис. 4.1

Для расчета вероятности успешности бурения продуктивных скважин вызовите Excel. Щелкните мышью на кнопке $\mathbf{f}_k$ (вставка функции). На экране компьютера высветиться окно **Мастер функций - шаг 1 из 2** (см .рис 4.2) . Выберите в списке **Категория Статистические** . в списке **Функции БИНОМРАСП**. Нажмите кнопку **OK**. На экране появится окно (рис 4.1).

Наберите (данные по примеру бурения) .

Число s = 0
Испытания = 5
Вероятность s = 0.1
Интегральный = 0

В окне (рис.4.1) высвечивается значение вероятности о продуктивности скважин (0,59049).
(При нажатии кнопки **OK** результат вносится в текущую ячейку текущего листа).

Задание 1. Просчитайте вероятности при:

1) n = 5; r = 1; p = 0,10; 2) n = 5; r = 5; p = 0,10;

Описательная статистика

По данным скважин глубокого бурения (столбец “Данные бурения”,табл 4.1./11/), с использованием **MS EXCEL** просчитать статистики ,характеризующие вариационный ряд.

Для этого:

1. Запустите **EXCEL**
2. В ячейку **A1** текущего листа введите заголовок таблицы (см. табл 4.1)
3. В ячейки **A2 : E2** – заголовки в соответсвии с заголовками табл. 4.1

(в **A2** - №, **B2** – Номер…, **C2** – Данные…, **D2** – Данные МОВ, **E2** – погрешность)

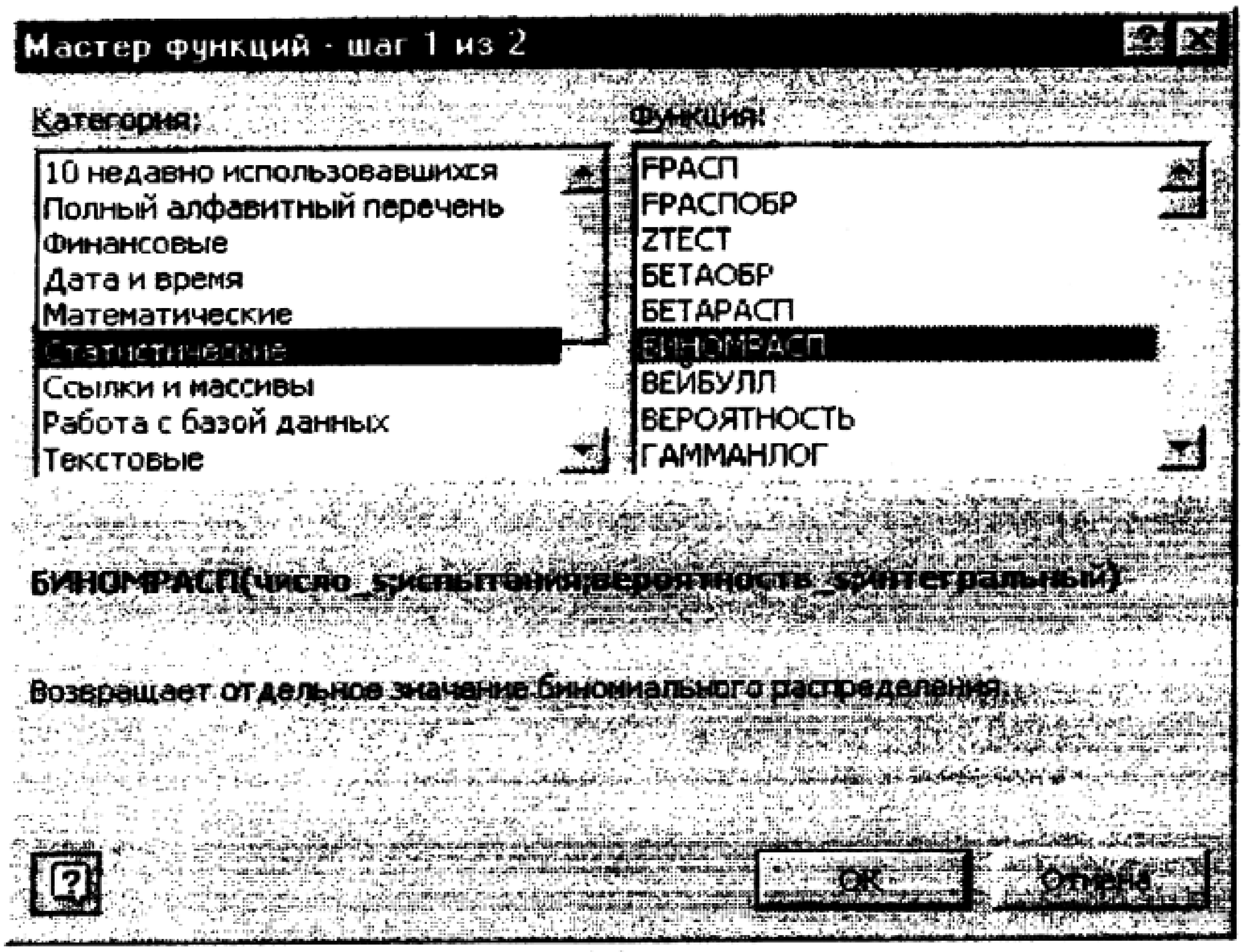


Рис.4.2

Таблица 4.1

Глубины залегания (в метрах) подошвы соленосного комплекса по данным глубокого бурения и сейсморазведки (МОВ).

№	Номер скважины	Данные бурения H_m	Данные МОВ * S_m	Погрешность, м ε_m
1	П-13	2300	2350	-50
2	Г-1	2395	2000	395
3	П- 11	3103	3050	53
4	П-38	2861	2640	221
5	П-88	3750	3400	350
6	Г-90	3733	3790	-57
7	Г-91	3856	3580	276
8	Г-93	3762	3400	362
9	Г-4	4384	3870	514
10	П-15	4220	3920	300
11	П-1	4687	4000	687
12	Г-4-1	5390	4600	790
13	Г-3	5390	4600	790
14	Г-7	925	950	-25
15	СГ-2	4880	4840	40
16	Г-1-1	3140	3290	-150
17	Г-2	3670	3370	300
18	Г-10	3662	3320	342
19	П-89	3706	3600	106
20	Г-1-2	3920	3400	520
21	Г-1-3	4818	4500	318

4. Столбцы **A:D** заполнить данными из табл.4.1

5.Столбец **E** рассчитать по формуле ε_m=**H**$_m$-**S**$_m$ (ε_m,**H**$_m$,**S**$_m$ измеряются в метрах).

Для этого в ячейку **E3** занесите формулу **=C3 –D3** и скопируйте в ячейки **E4:E23**.

(Столбец заполнится значениями ε_m.)

* МОВ – метод отраженных волн (сейсмический метод).

6. В меню EXCEL выбераем **Сервис – Анализ данных – Описательная статистика** . Открываем окно с помощью кнопки **OK**. (рис. 4.3)

7. В вызванном диалоговом окне **Описательная статистика** определяем интервал входных данных и область размещения результатов расчета (рис. 4.3).

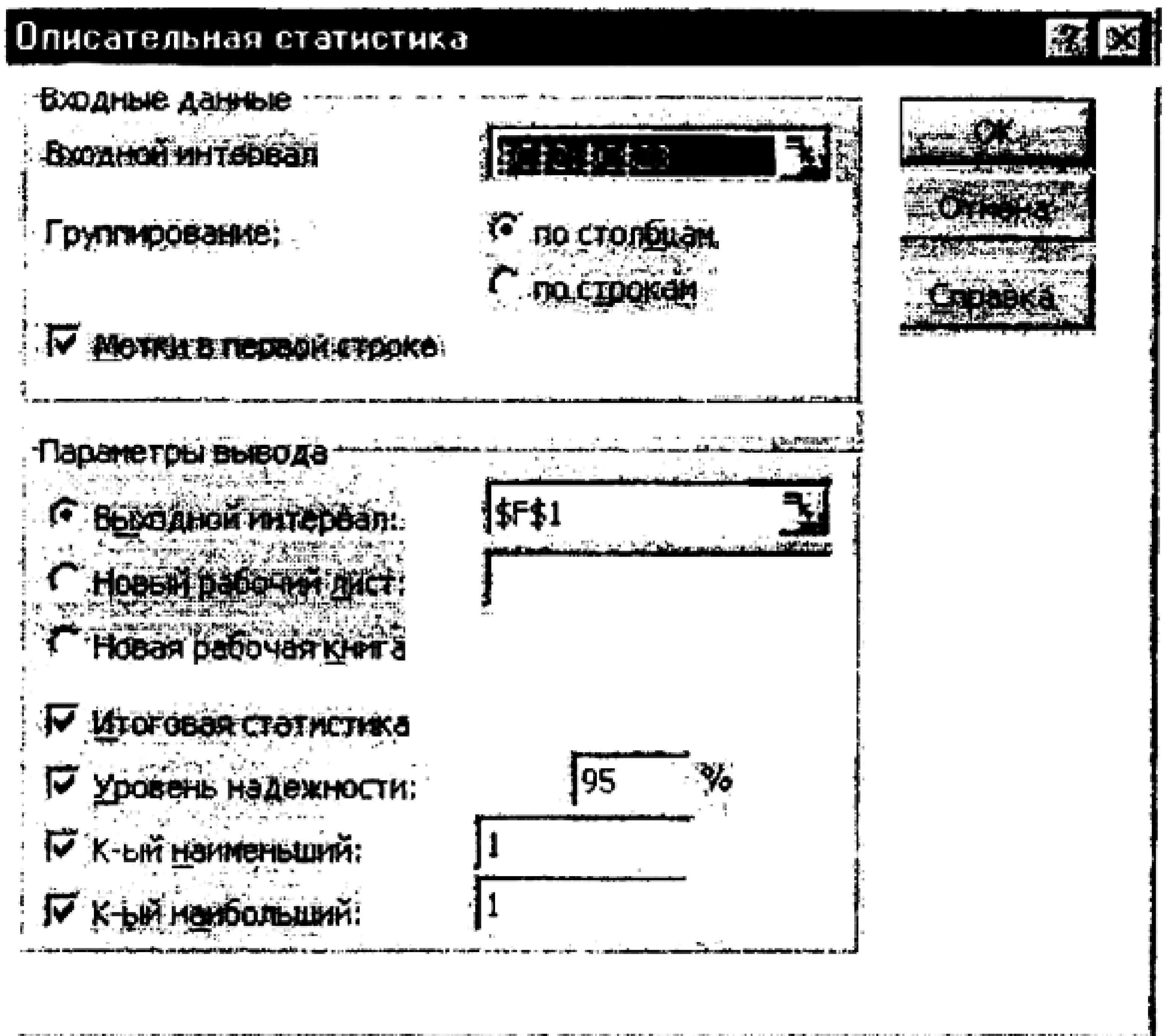


Рис 4.3

8 . Щелчок на кнопке ОК приводит к размещению в заданной области результатов. Результат определения описательных статистик для столбца “Данные бурения” на рис 4.4.

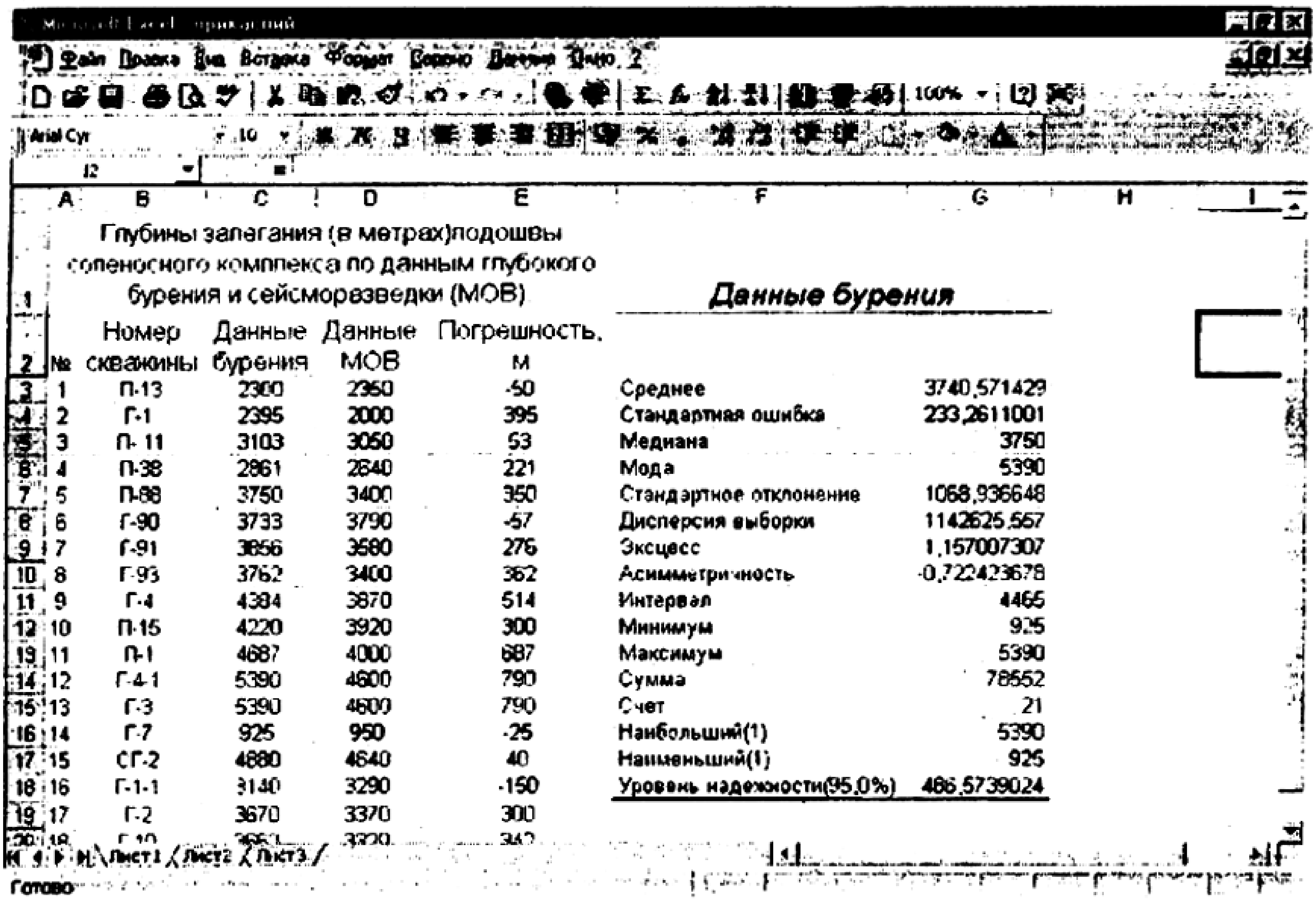
Уровень надежности (рис .4.4) есть ни что иное, как доверительная граница среднего с заданной вероятностью (часть 3, раздел 3.5).

В нашем примере 95% скважин глубокого бурения будут иметь глубину 3740,57 ± 486,57 метров (рис.4.4) до подошвы соленосного комплекса вне зависимости от места расположения скважины в регионе бурения.

Задание 2. Проведите расчеты статистик вариационного ряда **Данные МОВ** (табл. 4.1) аналогично расчетам ,проделанным для ряда **Данные бурения** .

Гипотезы, критерии

По данным табл.4.1 (переменные **“Данные бурения”** и **“Данные МОВ”**) провести сравнительный анализ с использованием **EXCEL** по **Z – критерию, t – критерию, F – критерию.**



Глубины залегания (в метрах) подошвы соленосного комплекса по данным глубокого бурения и сейсморазведки (МОВ)

№	Номер скважины	Данные бурения	Данные МОВ	Погрешность, м
1	П-13	2300	2360	-60
2	Г-1	2395	2000	395
3	П-11	3103	3050	53
4	П-38	2861	2640	221
5	П-88	3750	3400	350
6	Г-90	3733	3790	-57
7	Г-91	3856	3580	276
8	Г-93	3762	3400	362
9	Г-4	4384	3870	514
10	П-15	4220	3920	300
11	П-1	4687	4000	687
12	Г-4-1	5390	4600	790
13	Г-3	5390	4600	790
14	Г-7	925	950	-25
15	СГ-2	4880	4640	40
16	Г-1-1	3140	3290	-150
17	Г-2	3670	3370	300

Данные бурения

Среднее	3740,571429
Стандартная ошибка	233,2611001
Медиана	3750
Мода	5390
Стандартное отклонение	1068,936648
Дисперсия выборки	1142625,567
Эксцесс	1,157007307
Асимметричность	-0,722423678
Интервал	4465
Минимум	925
Максимум	5390
Сумма	78552
Счет	21
Наибольший(1)	5390
Наименьший(1)	925
Уровень надежности(95,0%)	486,5739024

Рис. 4.4

Z - критерий

1. Запустите **EXCEL**

2.В меню выберите **Сервис – Анализ данных – Двухвыборочный Z - тест для средних** (рис 4.5). Открываем окно с помощью кнопки **OK**.

3.В вызванном диалоговом окне (рис. 4.6) определяем интервалы входных данных и область (или лист, или книгу) размещения результатов расчетов.

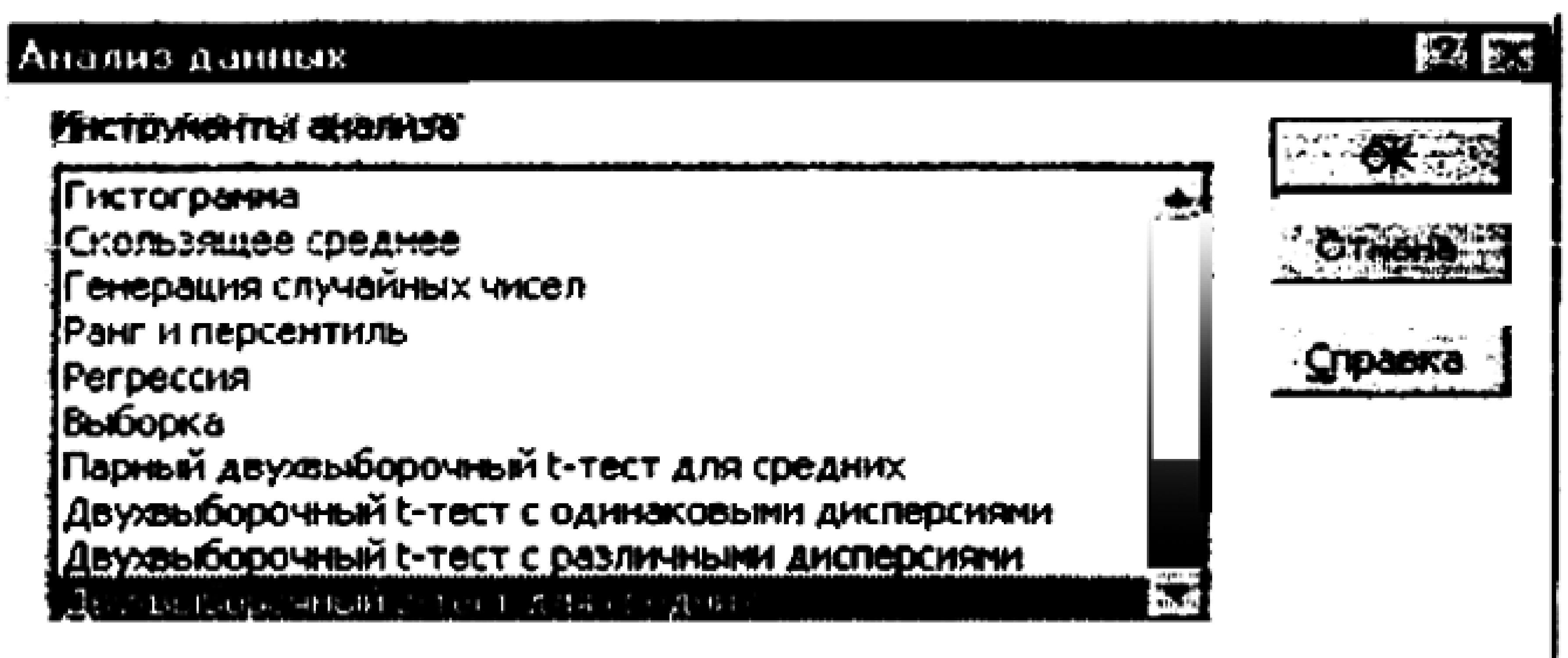


Рис. 4.5

Дисперсии переменных получены с использованием результатов описательной статистики (рис.4.4).

4. Щелчок на кнопке **OK** . Результат расчета на рис. 4.7

При интерпретации результатов счета воспользуйтесь раделами 3.2 ,3.3, 3.4, 3.5 третьей части данного пособия.

В просчитаном примере вычисленное значение (0,940) не превышает критического (1.959), из чего делаем вывод , что средние значения анализируемых переменных равны между собой при α =0.05.

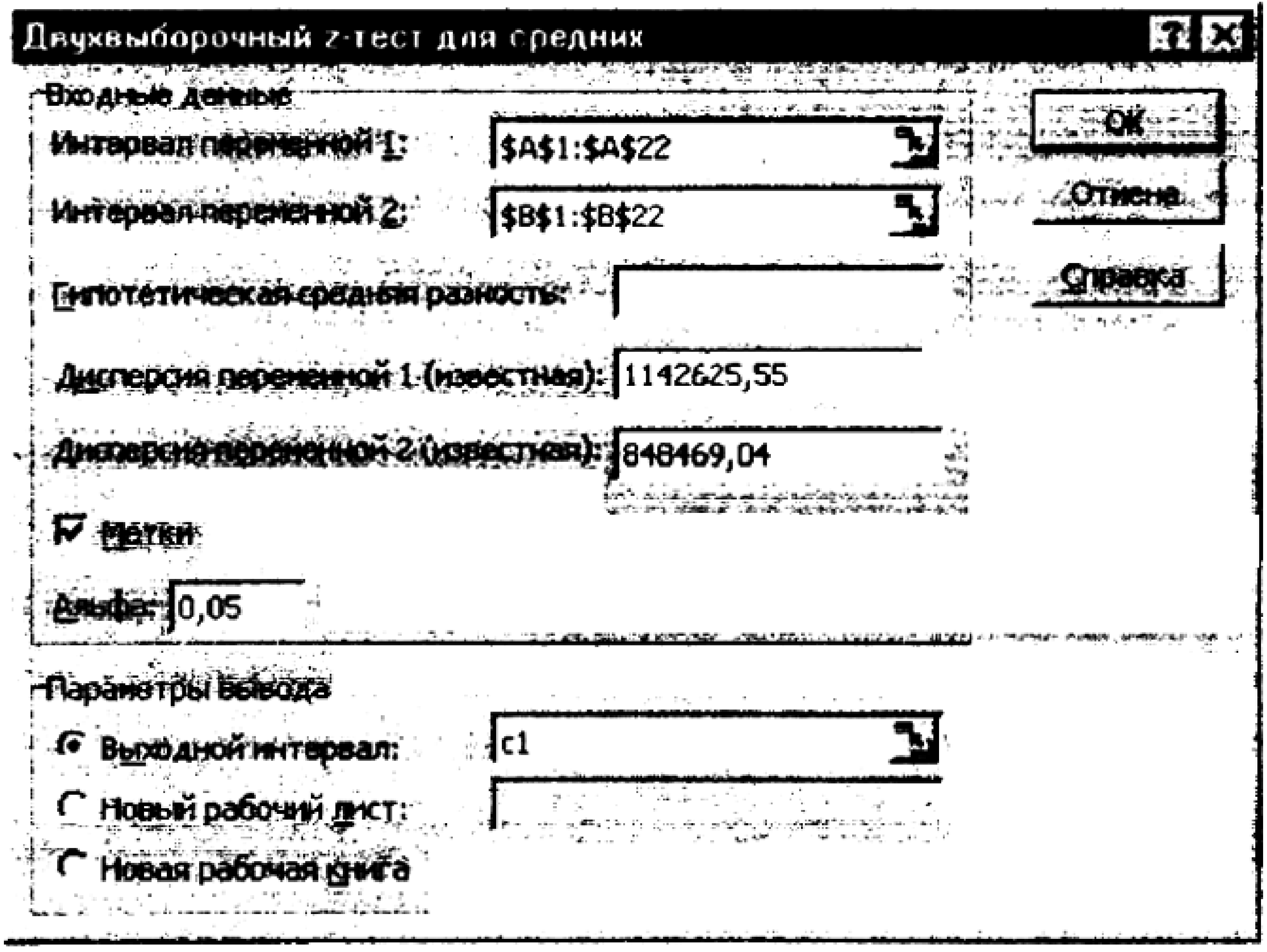


Рис 4.6

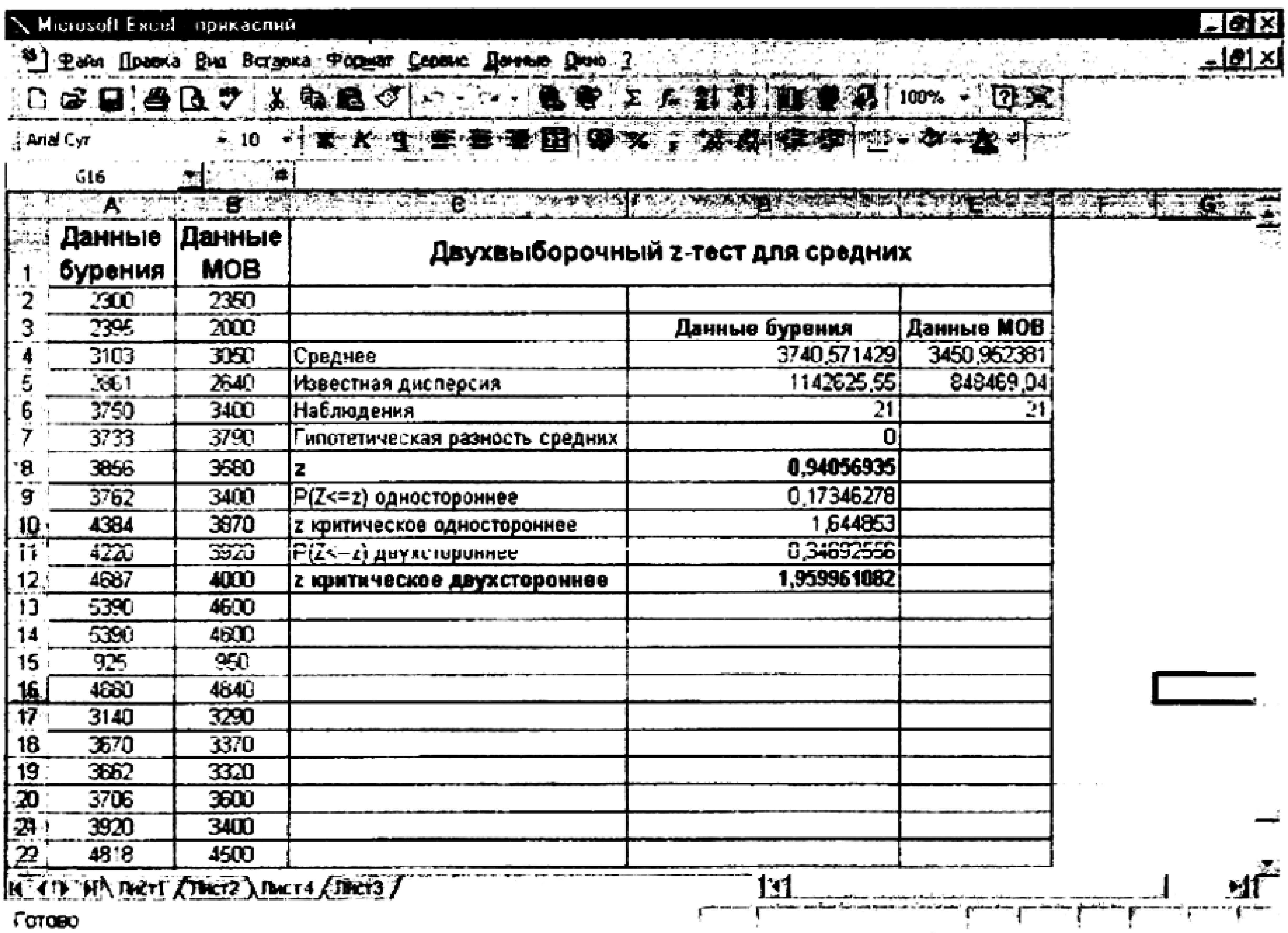


Рис. 4.7

Задание 3 Проведите сравнительный анализ переменных ε_m и H_m табл.4.1) по **Z – критерию**.

t- критерий

1.Запустите Excel.

2.В меню выберите **Сервис – Анализ данных – Двухвыборочный t – тест с одинаковыми дисперсиями.** (Предварительно дисперсии были просчитаны и проанализированы по F – критерию.) Открываем окно с помощью кнопки **OK** .

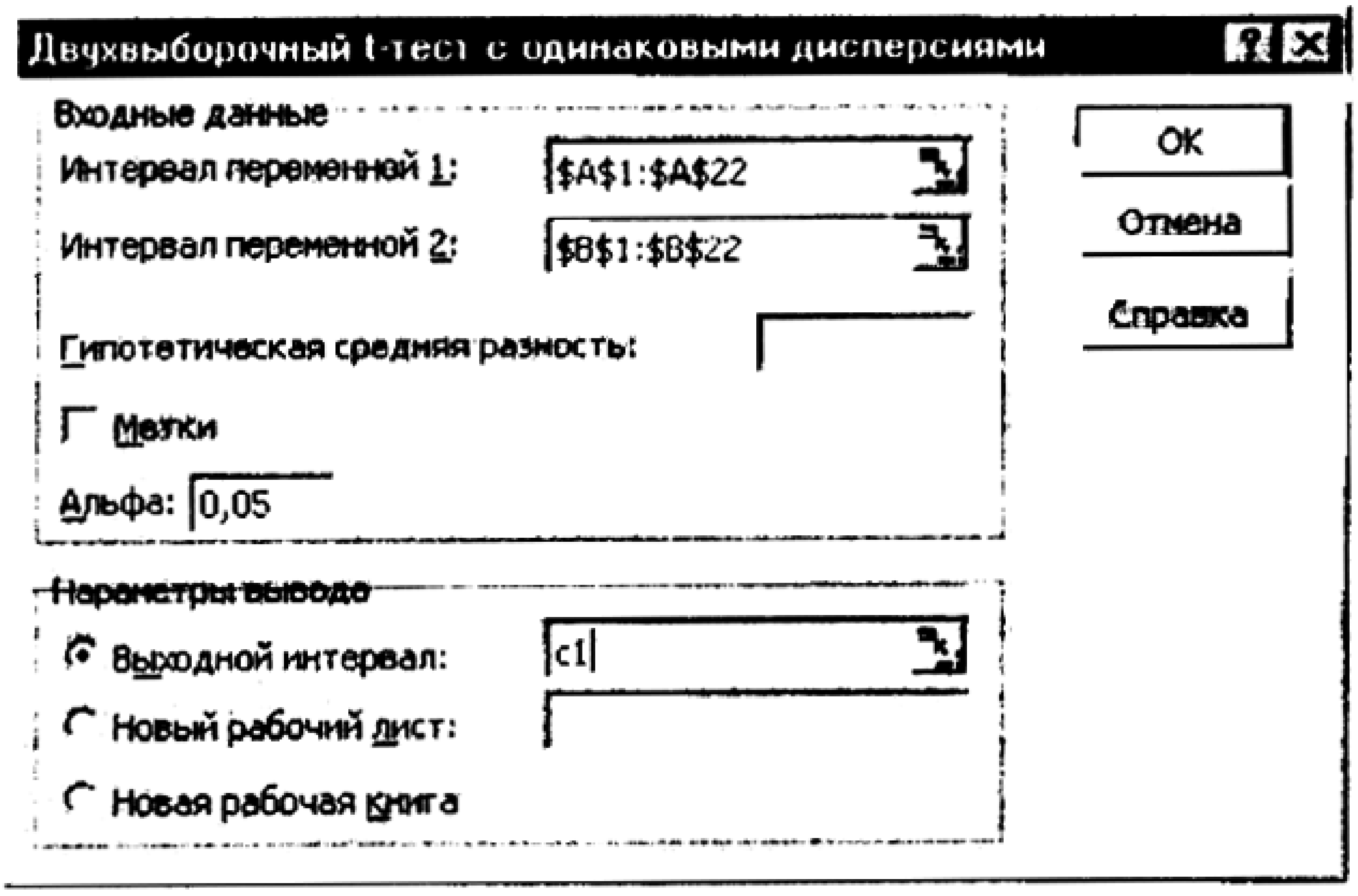
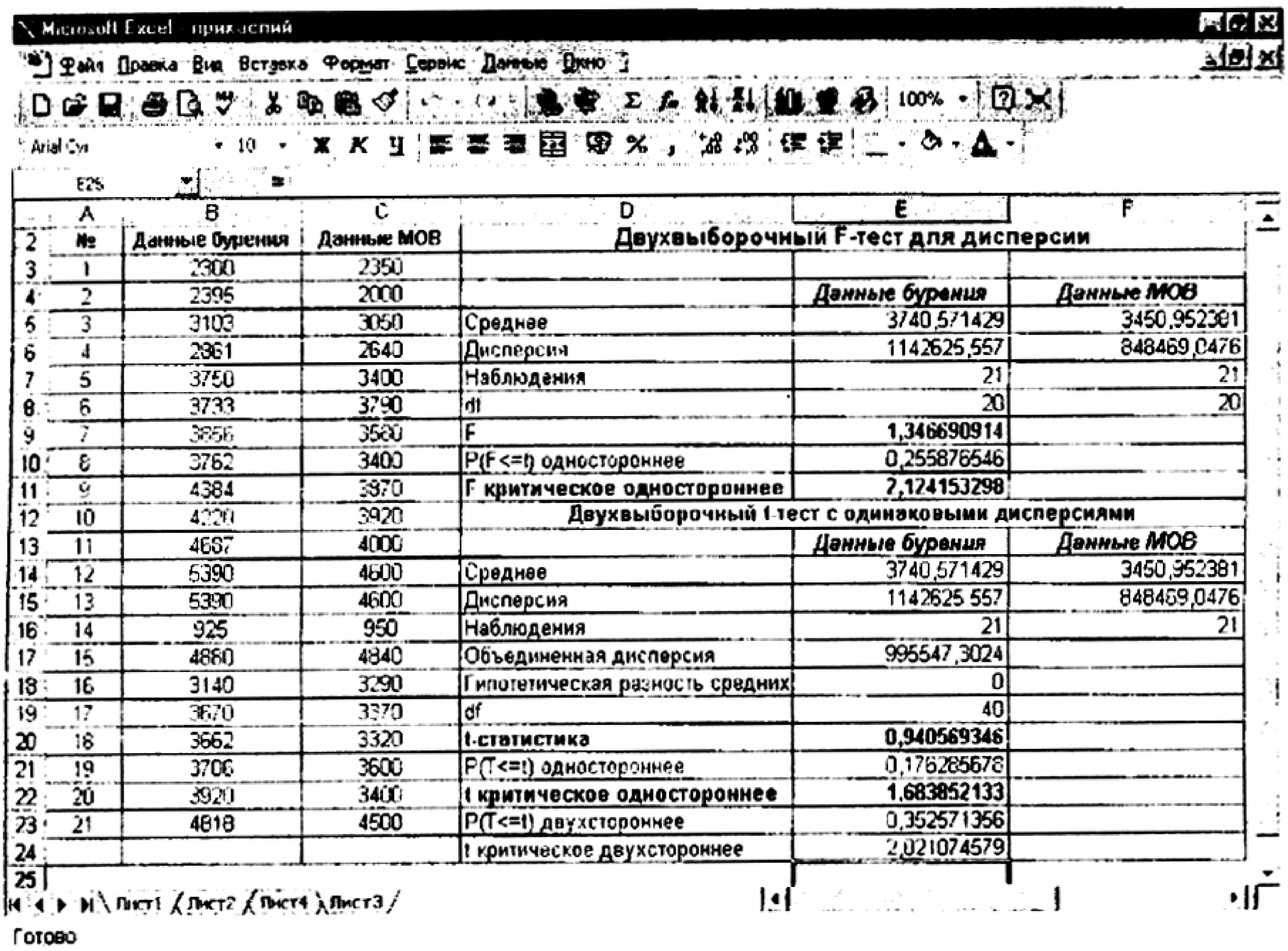


Рис. 4.8

3. В вызванном диалоговом окне (рис.4.8) определяем интервалы входных и выходных данных.
4. Щелчок на кнопке **OK** приводит к получению результата (рис. 4.9).



	A	B	C	D	E	F
2	№	Данные бурения	Данные МОВ	**Двухвыборочный F-тест для дисперсии**		
3	1	2300	2350			
4	2	2395	2000		*Данные бурения*	*Данные МОВ*
5	3	3103	3050	Среднее	3740,571429	3450,952381
6	4	2861	2640	Дисперсия	1142625,557	848469,0476
7	5	3750	3400	Наблюдения	21	21
8	6	3733	3790	df	20	20
9	7	3856	3580	F	**1,346690914**	
10	8	3762	3400	P(F<=f) одностороннее	0,255876546	
11	9	4384	3870	**F критическое одностороннее**	**2,124153298**	
12	10	4220	3920	**Двухвыборочный t-тест с одинаковыми дисперсиями**		
13	11	4667	4000		*Данные бурения*	*Данные МОВ*
14	12	5390	4600	Среднее	3740,571429	3450,952381
15	13	5390	4600	Дисперсия	1142625,557	848469,0476
16	14	925	950	Наблюдения	21	21
17	15	4860	4840	Объединенная дисперсия	995547,3024	
18	16	3140	3290	Гипотетическая разность средних	0	
19	17	3670	3370	df	40	
20	18	3662	3320	**t-статистика**	**0,940569346**	
21	19	3706	3600	P(T<=t) одностороннее	0,176285678	
22	20	3920	3400	**t критическое одностороннее**	**1,683852133**	
23	21	4818	4500	P(T<=t) двухстороннее	0,352571356	
24				t критическое двухстороннее	2,021074579	

Рис. 4.9

При интерпретации результатов по t – критерию (рис. 4.9) воспользуйтесь разделом 3.6 третьей части пособия. (В рассматриваемом примере расчитанное значение t – статистики равно 0.940 , t – критическое равно 1.683 и нет оснований предполагать , что изучаемые переменные имеют разные средние при α =0.05.).

Задание 4. Проведите сравнительный анализ переменных ε_m и H_m (табл.4.1) по **t- критерию.**

F - критерий

1.Запустите Excel.

2.В меню выберите **Сервис – Анализ данных – Двухвыборочный F– тест для дисперсии.** Открываем окно с помощью кнопки **OK** .

3. В вызванном диалоговом окне (рис.4.10) определяем интервалы входных и выходных данных.

4. Щелчок на кнопке **OK** приводит к получению результата (рис. 4.9).

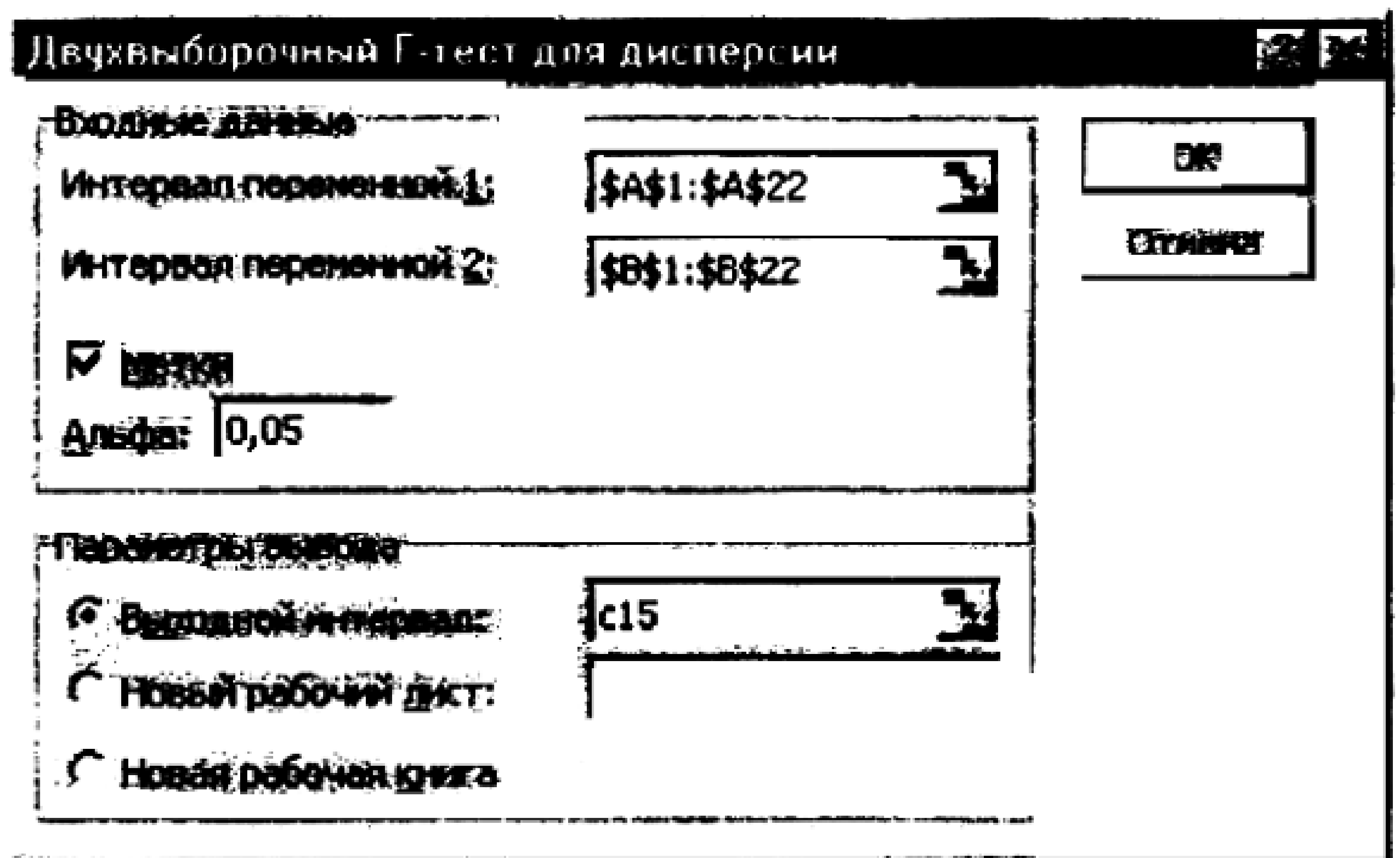


Рис. 4.10

При интерпретации результатов по F – критерию (рис. 4.9) воспользуйтесь разделом 3.8 третьей части пособия. (В рассматриваемом примере расчитанное значение F –критерия (1.346)
не превышает критическое значение равное 2.124 и нет оснований предполагать , что изучаемые переменные имеют разные дисперсии при α =0.05).

Задание 5. Проведите сравнительный анализ переменных ε_m и H_m (табл.4.1) по **F - критерию.**

Самостоятельное задание .

По данным табл.3.3 (брахиоды с плиты бассейна №1) и табл.3.4 (брахиоды с плиты бассейна №2) провести сравнительный анализ с использованием **EXCEL** по **Z – критерию, t – критерию, F – критерию.** (Обратите внимание на возможные расхождения в расчетах, полученных Вами и представленными в тексте пособия.)

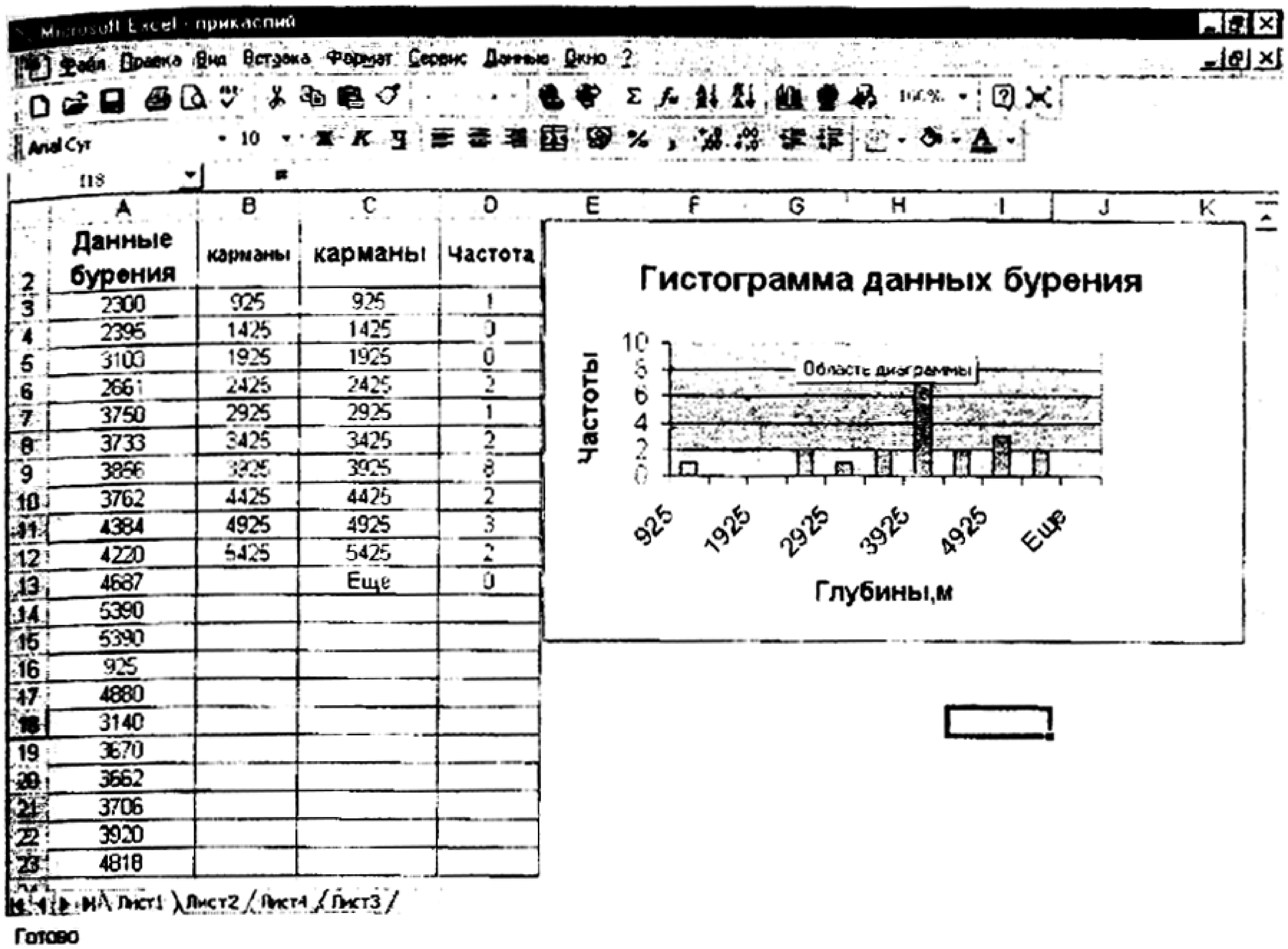


Рис. 4.11

Дополнительное самостоятельное задание .

Постройте гистограммы для переменных ε_m, H_m, S_m аналогично гистограмме ,показанной на рис. 4.11.(Оформление гистограммы на Ваш вкус).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Вентцель Е.С. Теория вероятностей.- М: Гос. изд-во физ-мат лит-ры 1962-564с.
2. Крамер Г. Математические методы статистики.- М: Мир 1975-648с.
3. Брант З. Статистические методы анализа наблюдений.- М: Мир 1975-312с.
4. Девис Дж.С. Статистический анализ данных в геологии.- М:Недра 1990.Т.1.319с; Т.2.427с.
5. Мантуров О.В., Солнцев Ю.К., Сорокин Ю.И. Толковый словарь математических терминов -М: Просвещение 1965.539с.
6. Гланц С. Медико-биологическая статистика.- М: Практика. 1999.459с.
7. Скрипченко Н.А. Анализ данных в MICROSOFT EXCEL.-Иркутск: Изд-во ИГТУ.
8. Дюк В. Обработка данных на ПК в примерах.- Питер Паблишин.1997.213с.
9. Колкот Э. Проверка значимости. –М: Статистика. 1978.128с.
10. Алферова М.А.,Михалевич И.М.,Рожкова Н.Ю. Сыклен А.Е. Примеры практической работы с Excel.Вып.2.-Иркутск, ИГИУВ,2001, -41 с.
11. Арабаджи М.С., Бакиров Э.А., Мильничук В.С.,Сенюков Р.В. Математические методы и ЭВМ в поисково – разведочных работах. М.Недра,1984.-264с.

ОГЛАВЛЕНИЕ