**Тема 3. Выборочный метод**

**Оглавление:**

3.1 Сплошное выборочное наблюдение

3.2 Статистические оценки

3.3 Оценка доли признака

3.4 Точечные оценки для средней и дисперсии генеральной совокупности

3.5 Интервальные оценки средней

*Цель:* ознакомить с методикой проведения выборочного обследования, определения ошибок выборки; распределению их на генеральную совокупность.

После изучения вы сможете: определять выборочные характеристики (средние, ошибки выборки) и распространить их на генеральную совокупность.

*Информационные источники:*

1. Статистика: Учебник/Под ред. И.И. Елисеевой. – М.: Крокус, 2008

2. Теория статистики: Учебник/Под ред. Г.П. Громыко. – М.: ИНФРА-М, 2000.

3. Галкина В.А. Статистика: Учебное пособие: М.: РГАЗУ,2002.

4. Курс теории статистики: Учебник/Под ред. В.Н. Салина, Э.Ю. Чурикова. – М.: Финансы и Статистика, 2006.

5. Статистика. Учебник/Л.П. Харченко, В.Г. Ионин, В.В. Глинский. -М.: ИНФРА-М, 2008.

Содержание темы: включает вопросы проведения и определения характеристик выборочного наблюдения. Основными понятиями являются виды отбора единиц совокупности; статистические оценки выборочной и генеральной совокупности.

**выборочное обследование генеральная совокупность**

**3.1 Сплошное выборочное наблюдение**

Статистическое наблюдение может быть *сплошным или выборочным.* Сплошное наблюдение предполагает наблюдение (измерение, исследование и т.д.) всех изучаемых объектов. Однако по ряду причин оно может оказаться принципиально неосуществимым или практически нецелесообразным. В таких случаях прибегают к наблюдению части изучаемых объектов и по его результатам делают выводы о свойствах всей совокупности. Такой метод наблюдения получил название *выборочного,* отобранная для изучения часть объектов называется *выборкой,* а вся исходная совокупность объектов — *генеральной совокупностью.*

Способ отбора элементов генеральной совокупности может быть *случайным* или *неслучайный.* При случайном отборе все элементы генеральной совокупности имеют равную вероятность попасть в выборку. Применение такого способа отбора позволяет положить в основу статистических выводов хорошо разработанные математиками вероятностные модели, закон больших чисел, методы изучения закономерностей случайных явлений.

Случайный отбор может производиться по схеме *возвращаемого (возвратная выборка)* или *невозвращаемого (безвозвратная выборка)* шара.

На практике выборка производится обычно как безвозвратная. Однако в теоретическом плане проще возвратная выборка, моделью которой служит схема повторных независимых испытаний. Поэтому в математической статистике, как правило, вначале подробно исследуется случай возвратной выборки, а затем указываются модификации статистических выводов при переходе к безвозвратному способу отбора.

Отличие этих выборок тем меньше, чем меньше отношение объема выборки к объему генеральной совокупности. Практически, если отношение составляет меньше 5—10%, этим отличием можно пренебречь и пользоваться более простыми соотношениями, предполагающими возвратную выборку.

**3.2 Статистические оценки**

Одна из важных задач математической статистики заключается в том, чтобы по данным случайной выборки оценить достаточно точно значения характеристик генерального распределения, как, например, долю признака, среднюю, дисперсию и т. д. Задачу об оценке можно разделить на две части: какую величину, подсчитанную по выборке, принять в качестве приближенного значения характеристики генерального распределения *(точечная оценка),* и в каком интервале вокруг этой величины будет заключена с заданной надежностью искомая характеристика *(интервальная оценка).*

Пусть генеральное распределение задается некоторой функцией *F(x,ξ1,…,ξк)*, где *ξ1,…*,*ξк* - его параметры. Например, если распределение задается двумя параметрами *ξ1* и *ξ2,* то *ξ1* обычно характеризует среднюю, а *ξ2-* дисперсию (или среднее квадратическое отклонение) генерального распределения.

Случайный отбор позволяет выборку объема *п* рассматривать как *п* повторных испытаний. Результат каждого испытания (*j-*го единичного отбора) есть случайная величина *Хj,* а вся выборка — совокупность *п* случайных величин {*Х1*, … *Хj,* ..., *Хп*}Любая конкретная выборка *(х1,* ..., *хi,* ..., *хп)* есть реализация этой совокупности случайных величин.

Для оценки неизвестного параметра *ξ* генеральной совокупности введем некоторую величину θ, вычисляемую по результатам выборки, т. е.

θ = θ (*X1,* ..., *Хj,* ..., *Хп),*

называемую *статистикой.*

Так, если для оценки генеральной средней *ξ =* выбрана статистика   
θ *= Х\** — выборочная средняя, то ее значения могут быть подсчитаны по результатам выборки как



Если для оценки генеральной дисперсии *D* выбрана статистика   
θ =D\* — выборочная дисперсия, то ее значения могут быть рассчитаны по формуле



Статистика θ есть случайная величина. В ряде случаев можно найти ее распределение.

Статистическая оценка должна быть возможно более точной. С этой целью к статистике θ предъявляются требования:

1. *состоятельности,*
2. *несмещенности,*
3. *эффективности.*

1) Свойство *состоятельности* означает, что распределение статистики θ с ростом объема выборки *п* концентрируется в сколь угодно малое окрестности параметра ξ (статистика θ стремится по вероятности к оцениваемому параметру ξ). Свойство состоятельности выражается предельным равенством: для любого столь угодно малого положительного числа ε

(1.9.1)



Свойство состоятельности может быть выражено двумя более жесткими требованиями, которые являются достаточными условиями состоятельности и которые легче поддаются практической проверке:

и (1.9.2)



2) Свойство *несмещенности* означает, что при любом конечном объеме выборки *п*центр рассеяния статистики θ (математическое ожидание случайной величины θ) совпадает со значением оцениваемого параметра генеральной совокупности:

М(θ) = ξ — для любого *п.* (1.9.3)

|  |
| --- |
|  |
| Рис. 1.9.1. Иллюстрация свойств состоятельности |

Естественно, что при заданном конечном объеме выборки *п* из различных возможных статистик для оценки параметра ξследует выбрать ту статистику, которая, являясь несмещенной, обладает в то же время минимальным рассеянием, т.е. имеет *минимальную дисперсию.* Последнее свойство получило название *эффективности.*

|  |
| --- |
|  |
| Рис. 1.9.2. Сравнение свойств трех статистик |

На рис. 1.9.2 показаны кривые распределения трех статистик. Из них θ и θ' — несмещенные и потому для построения оценки предпочтение должно быть отдано статистике θ' с меньшей дисперсией. Статистика θ" обладает еще меньшей дисперсией, однако она менее пригодна в качестве оценки, так как ее центр рассеяния смещен относительно параметра ξ`.

Статистику θ, принимающую для данной выборки определенное числовое значение, будем называть *точечной оценкой* параметра ξ и обозначать той же буквой, что и оцениваемый параметр, помечая ее звездочкой.

Для построения точечных оценок чаще всего применяют *метод аналогии,* т. е. для оценки параметров генерального распределения выбираются аналогичные параметры (характеристики) выборочного распределения.

Так, для оценки доли признака в генеральной совокупности p=M / N, генеральной средней и генеральной дисперсии



выбираются статистики (соответственно):

выборочная доля *р\*=*,



выборочная средняя



и выборочная дисперсия



При этом в результате дальнейшей проверки устанавливается, что первые две обладают свойством несмещенности, а последняя будет обладать этим свойством, если ее умножить на корректирующий множитель



Условия (1.9.2) и (1.9.3) позволяют для конечного *n* записать лишь приближенное равенство:

ξ≈ ξ\* (1.9.4)

Так как выборка носит случайный характер, то для различных возможных выборок случайная величина ξ\*может принимать различные значения. Поэтому возникает задача дополнить точечную оценку информацией о возможной ее погрешности, т. е. оценить ошибку выборки

δ= ξ - ξ\*

Пусть плотность распределения ξ\* изображена на рис. 1.9.3.

|  |
| --- |
|  |
| Рис. 1.9.3. Доверительные границы |

Выберем интервал (ξ – ε1, ξ +ε2), в котором с достаточно близкой   
к 1 вероятностью будет заключена величина ξ\*, т. е.

P(-ε1 <ξ - ξ\* <ε2) = l – α (1.9.5\*)

где α - величина, близкая к нулю.

Это означает, что в большинстве выборок (доля которых составляет   
1— α*)* ошибка выборки попадет в интервал (-ε1, ε2), и лишь в относительно малом числе выборок (доля которых равна α*)* ошибка δ выйдет за пределы интервала (-ε1, ε2 ). Поскольку производится одна выборка, то с практической достоверностью (т.е. с вероятностью 1 − α)можно полагать, что ее ошибка попадет в данный интервал, и, наоборот, практически невозможно (т. е. с вероятностью α),что она выйдет за границы интервала.

Но если ε1<ξ - ξ\* <ε2, то ξ\* - ε1< ξ< ξ\*+ ε 2, и равенство (1.9.5\*) запишется в виде:

P(ξ\* - ε1 <ξ <ξ\* +ε2) = l − α (1.9.5)

В силу изложенного

* интервал (ξ\* - ε1, ξ\*+ε2) называется *доверительным интервалом,*
* числа ξ\*- ε1, ξ\*+ε2 - *доверительными границами,*
* вероятность Р=1—α *- доверительной вероятностью* и
* α- *уровнем значимости (существенности)*

Доверительный интервал дополняет точечную оценку *ξ*\* оценкой ошибки выборки, или интервальной оценкой параметра α.

Если для точечной оценки необходимо знать лишь выражение для *ξ*\* как функцию данных выборки, то для построения доверительного интервала необходимо знать также закон распределения *ξ*\*, с помощью которого рассчитывается вероятность (1.9.5).

Часто при симметричном характере распределения случайной величины *ξ*\* относительно *ξ* можно и доверительный интервал рассматривать как симметричный относительно *ξ*. В таком случае уравнение (1.9.5) может быть заменено на более простое:

P(ξ\* - ε <ξ <ξ\* +ε) = P (│ξ - ξ\*│<ε) = l – α (1.9.6)

Величина ε называется *предельной ошибкой выборки.*

С интервальной оценкой связано решение трех типов задач:

1) определение доверительного интервала по заданной доверительной вероятности *Р= 1 – α* и объему выборки *п*;

2) определение доверительной вероятности по заданному доверительному интервалу и объему выборки;

3) определение необходимого объема выборки *п* по заданным доверительной вероятности и доверительному интервалу.

**3.3 Оценка доли признака**

Для точечной оценки доли признака в генеральной совокупности *(р)* естественно взять выборочную долю

*р*\**=*



где *n* — объем выборки,

*т* — количество единиц в выборке, обладающих данным признаком.

Можно доказать, что эта оценка является *состоятельной, несмещенной, эффективной.*

Вопрос об интервальной оценке рассмотрим сначала для случая *возвратной выборки.*

При такой организации выборки случайная величина *p\*,* как известно из теории вероятностей, имеет биномиальный закон распределения. Расчет доверительного интервала с применением формулы биномиального закона связан с определенными вычислительными трудностями. Однако при достаточно большом объеме выборки (примерно *n* ≥ 20, *пр* ≥ 10*)* биномиальное распределение хорошо аппроксимируется нормальным распределением с параметрами

*М (p\*) = p*;

σ*(p\*)* =



Следовательно, случайная величина имеет стандартное нормальное распределение (с параметрами M(z)=0; σ(z)=1).



Задавшись определенной вероятностью Р=1— α, имеем:

2**Ф**(zα)=1- α (1.9.7)



где **Ф**(zα)= — интегральная функция Лапласа, значения которой для различных значений *z* рассчитаны и приводятся в специальных таблицах.



Равенство (1.9.7) эквивалентно равенству:

*P {│p\*- p │<z1 ·* σ( *p*\*)} = 2**Ф**(zα*)* (1.9.7')

Таким образом, предельная ошибка выборки εα определяется из равенства:

(1.9.8)



Применение этой формулы затрудняется тем, что в нее входит неизвестный параметр *р* — генеральная доля. Однако при большом *п* можно заменить неизвестный параметр *р* его точечной оценкой *р\*.* Тогда получим:

(1.9.9)



Приведенные выше формулы связывают между собой, в конечном счете, три величины: доверительную вероятность Р=1−α, предельную ошибку выборки ε и объем выборки *п.*

Вкаждой конкретной задаче две из этих величин задаются и определяется третья величина. Таким образом, мы имеем следующие три типа задач:

1. Даны *п* и *Р*, определить ε.
2. Даны *п* и ε, определить *Р*.

III. Даны *Р* и ε, определить *п*

Первые два типа задач связаны с анализом результатов уже произведенной выборки объема *п,* следовательно, и с найденной точечной оценкой *р\*.*

Задачи третьего типа должны решаться до проведения выборки. По заданной доверительной вероятности *P* мы можем определить величину *z* (по таблице интегральной функции Лапласа). Из (1.9.9) получаем:

(1.9.10)



Но в (1.9.10) входит величина *р\**, получаемая в результате выборки, а речь идет об определении *п* до осуществления выборки.

Поскольку *р\** неизвестно, то определяем из этого равенства, при каком значении *р\** величина *п* будет максимальной. Используя обычный метод следования функции на максимум, получаем:



откуда *р\**=½

Следовательно,

(1.9.11)



Выборка такого объема наверняка обеспечит заданные надежность и точность.

Рассмотрим примеры на каждый из трех типов задач. Исследуется вопрос о доле поврежденных клубней картофеля после механической уборки.

**Пример 1.9.1** Произведена случайная выборка объемом.n=200 деталей. Из них поврежденных оказалось 40. Определить *с* вероятностью 0,95 доверительный интервал для доли поврежденных деталей генеральной совокупности.

Рассчитываем выборочную долю:

*р\* = m / n* = 40 / 200 = 0.20

По заданной доверительной вероятности

*Р = 1 – α =* 2**Ф**(zα) = 0.95

находим по таблице интегральной функции Лапласа соответствующее значение *zα* =1,96. Применяем формулу (1.9.9):



Таким образом, доверительный интервал для генеральном доли *р:*

0,20-0,06<p<0,20+0,06, или 0,14<p<0,26

**Пример 1.9.2.** По результатам той же выборки определить вероятность того, что ошибка выборки не превысит 0,03.

Имеем:



Отсюда:



По таблице интегральной функции Лапласа находим соответствующую доверительную вероятность *Р =* 2**Ф***(zа* )=0,71.

**Пример 1.9.3.** До проведения выборки необходимо ответить на вопрос: какой объем выборки обеспечит с вероятностью0,95 ошибку выборзки не более, чем 0,02?

Применяем формулу (1.9.11):



Следует заметить, что требуемые надежность и точность может обеспечить в нашей задаче и выборка меньшего объема.Если до проведения выборкиу нас есть приближенная оценка хотя бы максимальной величины *р\**,то мыможем применить формулу (1.9.10) и получить меньшее значение необходимого объема выборки *п.*

В случае *безвозвратной выборки* случайная величина *р\*,* как доказываетсяв теории вероятностей, имеет так называемое гипергеометрическое распределение. Ее математическое ожидание,как и *в* случае возвратнойвыборки, равно генеральной доле: *М*(*р\**)*=р,* а среднее квадратическоеотклонение вычисляется но формуле:

(1.9.12)



где *N* — объем генеральной совокупности

Придостаточно большом объеме выборки гипергеометрическоераспределение также хорошо аппроксимируетсянормальным распределением с указанными параметрами *M*(*p*\*) и *σ*(*p*\*), поэтому дальнейший ход решения задач аналогичен рассмотренному выше случаю возвратной выборки.

Формула для предельной выборки принимает вид

(1.9.13)



При решении задач III типа из (1.9.13) получаем:

(1.9.14)



Соответственно изменится и формула для *nmax* :

(1.9.15)



Если объем выборочной совокупности *n* составляет незначительную долю по отношению к объему генеральной совокупности *N*, то величина в формуле (1.9.12) ближе к 1, можно пренебречь различием формул (1.9.9) и (1.9.13) и пользоваться более простыми соотношениями для возвратной выборки, даже если фактически выборка производится как безвозвратная.



В заключение раздела необходимо отметить что в статистике используется понятие *средней ошибки выборки*, которая определяется как среднее квадратическое отклонение соответствующей выборочной характеристики. Нетрудно видеть, что формула для средней ошибки выборки является частным случаем формулы предельной ошибки выборки при z=1.

**3.4 Точечные оценки для средней и дисперсии генеральной совокупности**

Обозначим через и σ2 среднюю и дисперсию генеральной совокупности.



*Возвратная выборка объема n* может рассматриваться как совокупность n независимых случайных величин *Xj*, имеющих одно и то же распределение, совпадающее с генеральным, для которых, следовательно:

*M*(*Xj*) *= ; D*(*Xj*)= σ2



Для точечной оценки генеральной средней естественно использовать статистику ⎯ среднюю. Используя свойства математического ожидания и дисперсии, получим:



(1.9.16)



(1.9.17)



Нетрудно видеть, что статистика θ ⎯ *X\** является состоятельной, несмещенной и эффективной оценкой параметра .



Для точечной оценки генеральной дисперсии воспользуемся статистикой *—* выборочной дисперсией. Однако при ближайшем рассмотрении оказывается, что



(1.9.18)



Таким образом, статистика θ *= D\** является смещеннойоценкой для генеральной дисперсии σ2. Однако смещенность легко устраняется путем введения корректирующего множителя .Статистика



(1.9.19)



(так называемая «исправленная» выборочная дисперсия) является несмещенной оценкой генеральной дисперсии σ2 и используется для ее точечной оценки.

Заметим, что при большом *п* отношение и потому значение s2≈*D\**



В случае *безвозвратной выборки можно* показать, что точечная оценка средней будет той же (т. е. \**),* а точечная оценка дисперсии должна быть заменена на:



(1.9.20)



где N — объем генеральной совокупности

В случае безвозвратной выборки изменится и выражение для *D*(\*)*,* которое потребуется для построения доверительного интервала при оценке средней:



(1.9.21)



При относительно небольшом объеме выборки и



**3.5 Интервальные оценки средней**

При изложении данного вопроса будем различать случаи *больших* и *малых* выборок. При этом оба случая сначала рассмотрим в более простой, с теоретической точки зрения, ситуации *возвратной* (повторной) выборки.

***3.5.1 Большая выборка***

Если объем выборки достаточно большой (практически, начиная с *п* > 20—30), то распределение выборочной средней *,* согласно центральной предельной теореме, независимо от характера генерального распределения приближается к нормальному распределению с параметрами



*М*()*=* и *)*



где  *—* генеральная средняя,



σ— генеральное среднее квадратическое отклонение,

*п* — объем выборки.

Таким образом, величина



распределена по стандартному нормальному закону (с математическим ожиданием *M*(*z*) *=* 0и средним квадратическим отклонением σ*(z) =* 1).

Задавшись доверительной вероятностью *Р* = 1 — α*,* определяем из равенства 2Ф(z) = 1 — α соответствующее значение *za* (используем при этом таблицу интегральной функции Лапласа). Тогда с вероятностью *Р* = 1 — α выполняется неравенство:

(1.9.22)



которое эквивалентно неравенству:

(1.9.23)



Величина называется предельной ошибкой выборки.



Таким образом, мы имеем доверительный интервал для генеральной средней:

( ; )



Наоборот, если задана предельная ошибка ε , а требуется определить вероятность *Р,* то схема решения задачи следующая:

ε→z=→Ф(*z*)→*P*=2Ф(*z*) (1.9.24)



Наконец, определение объема выборки *п* по данным *Р* и ε производится по следующей схеме:

*P*=2Ф(*z*) →*z*→*n*= (1.9.25)



**Пример 1.9.4.**  Взвешивание 50 случайно отобранных коробок печенья дало =1200г. Определить с вероятностью *Р* = 0,95 доверительные границы для среднего веса коробки печенья в генеральной совокупности, если есть основания полагать, что генеральная дисперсия σ*2 =* 11664.



***Решение:***

Дано: *n*=50; *=*1200; σ2 =11664 (*=* 108); Р = 0,95.



Из равенства *Р* = 2Ф(*z*)=0,95 по таблице значений интегральной функции Лапласа находим z=1,96, откуда:

ε=(г)



Таким образом, получаем доверительный интервал:

1200 — 30 < < 1200 + 30.



**Пример 1.9.5** Определить, с какой доверительной вероятностью можно утверждать, что при данном объеме выборки (50 коробок) ошибка выборки не превысит 20 г.

***Решение:***

По величине ε=20 вычисляем , откуда по таблице Ф(*z*): *Р* = 2Ф(1,31)≈0,81



**Пример 1.9.6.** Определить необходимый объем выборки *n*, который с вероятностью 0,99 гарантировал бы ошибку выборки не более чем ε *=* 20г.

***Решение:***

Из *Р =* 2Ф(*z) =0,99* находим *z* = 2,58, откуда:

коробок



Предположение о том, что генеральная дисперсия σ***2***известна при неизвестной генеральной средней, на практике выполняется весьма редко. Чаще всего мы имеем лишь выборочные данные и можем дать лишь выборочную оценку *s2* неизвестной дисперсии σ***2****.*

Статистика

(1.9.26)



подчиняется закону распределения Стьюдента с *v = n*—1 степенями свободы. Однако при больших значениях параметра *v* (*v* ≥ 30) распределение Стьюдента практически совпадает с нормальным. Поэтому в случае больших выборок схема решения задач остается прежней, даже если вместо 'Неизве стного генерального среднего квадратического отклонения а используется его выборочная оценка *s.*

***3.5.2. Малая выборка***

Если генеральная совокупность подчинена нормальному закону распределения (что на практике имеет место очень часто), то выборочная средняя как средняя арифметическая *п* нормально распределенных случайных величин также имеет нормальный закон распределения. Таким образом, величина распределена по стандартному нормальному закону, и схема решения задач при известном генеральном среднем квадратическом отклонении σ остается прежней.



Если же генеральное среднее квадратическое отклонение *σ* неизвестно и приходится пользоваться его выборочной оценкой *s,* то используется статистика *t* (1.9.26), которая, как мы уже отмечали, подчинена закону распределения Стьюдента с *v = n*—1 степенями свободы. При *v* < 30 имеются значительные различия между распределением Стьюдента и нормальным распределением (тем более значительные, чем меньше *v*). Используя функцию распределения Стьюдента, мы можем записать равенство, аналогичное формуле Лапласа:

(1.9.27)



где *S(t, v*) — функция Стьюдента, значения которой для различных значений *t*и *v* подробно рассчитаны и представлены в специальных таблицах.

Выражение **(**1.9.27**)** эквивалентно выражению:

(1.9.28)



где



Решение задач с помощью этого равенства аналогично решению задач с использованием формулы Лапласа. Лишь определение *п* несколько усложняется из-за того, что оно входит также в параметр *v = n*—1.

Поэтому можно воспользоваться схемой последовательных приближений. Вначале производят оценку (*s*2) генеральной дисперсии. Затем находят *п1* по схеме (1.9.25), используя таблицу функции Лапласа и принимая *σ2 = s2-* По найденному *n1* и, соответственно, *v1 = n1* — 1 и заданному значению

*Р*=1—*α* определяют *t1* (по таблице распределения Стьюдента) и вычисляют и так далее.



Теперь можно снова повторить расчет по *v2 = n2* — 1 и т.д.

Итерация заканчивается, если окажется *ni* ≈ *ni-1*.

**Пример 1.9.7.** Для определения среднего заработка работника за день при соблюдении необходимых условий было отобрано 10 работников, заработок которых оказался равным (в руб.): 325; 337; 319; 330; 327; 328; 332; 320; 318; 334. Требуется определить с вероятностью 0,95 доверительный интервал для среднего заработка работников в генеральной совокупности, если есть основания полагать, что заработная плата в генеральной совокупности подчиняется нормальному закону определения.

***Решение:***

По данным выборки определяем среднюю и дисперсию. Получаем

;



Рассчитываем несмещенную оценку генеральной дисперсии



Предположение о нормальном характере генерального распределения позволяет нам использовать равенства (1.9.27) и (1.9.28). Обращаясь к таблице значений функции Стьюдента, по заданным *P = 2S(t, v*)=0,95 и *v* = *n*—1 = 10 – 1 = 9 находим *t* = 2,26.

Вычисляем предельную ошибку выборки ε=(кг).



Доверительный интервал для генеральной средней:

327—5<<327+5 или 322<<332.



**Пример 1.9.8.** Используя данные примера 1.9.7, определить объем выборки, необходимый для того, чтобы ошибка выборочной средней с вероятностью 0,95 не превышала 3 рубля.

**Решение.**

Мы имеем оценку генеральной дисперсии s2 = 42,4. Вначале находим *n1* по формуле (1.9.25), принимая σ*2 = s2* и определяя *z* по таблице функции Лапласа:



Теперь обращаемся к таблице функции Стьюдента и по *Р* = 0,95,

*v1* = *n1*—1 ≈ 17 находим значение *t1*=2,11.

Вычисляем



По *Р = 0,95* и *v2* = *n2*—1 = 21 – 1 = 20 находим *t*2 = 2,09.

Вычисляем



Поскольку *n3* ≈ *n2 ,* то необходимый объем выборки устанавливается 21 человек.

Еще раз отметим, что рассмотренные выше схемы решения задач для малых выборок справедливы только при предположении нормального характера генерального распределения. При отсутствии такого предположения распределения неизвестно, и выборочную среднюю можно использовать лишь как точечную оценку генеральной средней без оценки точности .приближенного равенства *,* т. е. без расчета доверительного интервала.



***3.5.3.******Безвозвратная выборка***

В случае безвозвратной выборки формула для среднего квадратического отклонения выборочной средней, согласно (2.21), примет вид:

(1.9.29)



Если генеральное среднее квадратическое отклонение σнеизвестно (наиболее реальная ситуация), то мы заменяем его точечной оценкой s', которая рассчитывается по формуле (1.9.20). В результате получим:

(1.9.30)



(*.s* — обычное «исправленное» среднее квадратическое отклонение

)



Во всем остальном ход решения задач как для случая больших выборок, так и для случая малых выборок остается прежним.

Корректирующий множитель при малой величине (например, для 1 или 5% выборок) близок к 1, и поэтому расчеты могут производиться как для возвратной выборки.

