Статистическое изучение модификационной изменчивости.

Построение вариационных рядов

Подготовка материала к математической обработке

Предметом изучения биометрии являются варьирующие (изменяющиеся) признаки у относительно однородной группы объектов, называемой совокупностью.

Различают совокупность генеральную и выборочную, или случайную, которую называют выборка.

Генеральной совокупностью может быть любая группа особей. Изучается же, как правило, часть членов генеральной совокупности, так называемая выборочная совокупность, или выборка.

Основное требование к выборке – ее репрезентативность, т.е. достоверное отражение генеральной совокупности. Достигается это случайным отбором объектов для формирования выборки. Объем генеральной совокупности принято обозначать буквой **N**, а выборочной – **n**. Если в выборку включено меньше 30 членов (n<30), ее называют малой, во всех других случаях – большой.

По характеру изменчивости различают изменчивость качественную и количественную. Количественная изменчивость может быть дискретной, т.е. прерывистой (она выражается только целыми числами: количество детей при рождении, число сосков у многоплодных животных и т.д.), и непрерывной, когда варианты могут принимать любое значение (вес, рост, размер обуви, объем крови в организме и т. д.). Величина признака отдельной особи, т.е. числовое его значение, называется вариантой и обозначается буквой **Х.** Обработка варьирующих показателей начинается с их группировки. Способ группировки зависит как от характера изменчивости, так и от объема выборки.

 Наиболее простым методом группировки, применяемым при любом характере изменчивости при небольших размерах выборки (n<30), является ранжирование.

***Пример***: вес новорожденных при рождении был равен (в кг):

3,3; 3,1; 3,2; 4,2; 3,4; 3,0; 3,8; 3,3; 3,2; 4,2.

Ранжированный ряд этой выборки будет выглядеть так:

3,0; 3,1; 3,2; 3,2; 3,3; 3,3; 3,4; 3,8; 4,2; 4,2.

Сущность ранжирования состоит в том, что варианты располагаются в строгом порядке по принципу их увеличения или уменьшения (т.е. по ранжиру). Минимальный вес новорожденных 3,0 кг, максимальный – 4,2 кг.

Для больших выборок (n>30) основным методом группировки является построение вариационного ряда.

 Вариационным рядом называется двойной ряд, отражающий распределение вариант по классам.

 При составлении вариационного ряда значения, которые принимает признак, называют классами (**W**), а количество вариант в классе – частотами (**p** или **f**). Сумма частот по всем классам должна равняться объему выборки (n). В математическом выражении это записывается так: Σp = n. Если признак имеет большой размах изменчивости, то в этом случае в один класс рекомендуется объединить варианты с несколькими, близкими между собой, значениями.

 Классы вариационного ряда в таком случае не будут совпадать со значениями, которые признак может принимать в процессе изменчивости, а будут характеризоваться несколькими показателями: началом класса (**Wн**), т.е. минимальным значением признака, концом класса (**Wк**), т.е. максимальным значением признака.

# Построение вариационного ряда

 Разберем на конкретном примере построение вариационного ряда.

***Пример***. При взвешивании 50 спортсменов получены следующие данные (в кг):

58 50 53 53 50 61 58 58 57 52

49 51 63 55 50 57 66 46 60 53

58 53 50 54 50 51 67 47 52 47

47 54 59 54 53 57 52 50 46 56

42 55 52 57 54 56 50 59 49 54

 Для составления вариационного ряда необходимо:

1. Найти в учетах данных максимальное (max) и минимальное (min) значения признака.

Разница между максимальным и минимальным значениями признака (варианта) – это размах изменчивости признака (lim = max–min).

1. Исходя из объема выборки и размаха изменчивости, выбрать оптимальное число классов (k) для проведения группировки.

 Число наблюдений Число классов

 40–60 6–10

 61–100 7–10

 101–200 9–12

 201–500 12–17

В нашем примере число измерений равняется 50. Значит , число классов должно быть в пределах 6–10. В этих пределах подбирать число классов следует таким образом, чтобы величина классового промежутка была удобной для подсчета и, желательно, оканчивалась на цифру 5 или 0.

3. На основании выбранного количества классов и размаха изменчивости признака установить величину классового промежутка (i), т.е. величину, на которую один класс должен отличаться от другого:

max = 67; min = 42; lim = 25; k = 8 (подобранное нами число классов = 8)



Началом первого класса обычно служит варианта с минимальным значением признака, концом первого класса – величина, равная началу первого класса, увеличенному на классовый промежуток (i). Конец последнего класса завершается максимальным значением варианты. Конец предыдущего и начало следующего классов не должны совпадать. Они должны отличаться или на целое число, или на десятые или сотые доли числа, в зависимости от величины изучаемого признака. Установленные для нашего примера границы классов заносятся в табл.1.

## Статистические показатели для характеристики совокупности

### Среднее значение признака

 Полученные при проведении обследования данные характеризуют каждую особь совокупности в отдельности. Нас же интересуют, в первую очередь, наиболее общие свойства этой совокупности. Чтобы их установить, данные обрабатывают статистически. Основная задача статистической обработки наблюдений – нахождение ряда показателей, характеризующих в обобщенном виде свойства данной совокупности.

 Одним из таких показателей является средняя арифметическая, характеризующая среднее значение признака.

### Средняя арифметическая

 Средняя арифметическая представляет собой как бы точку равновесия вариационного ряда, отклонения от которой в сторону увеличения или уменьшения признака взаимно уравновешиваются. Средняя арифметическая показывает, какую величину признака имели бы особи данной группы, если бы эта величина была у всех одинаковой.

 Простейший метод вычисления средней арифметической величины для небольшой выборки (n<30) – это простое суммирование, т.е. нахождение суммы вариант выборки и деление ее на объем выборки. Среднюю арифметическую обозначают **Хср** или **М**.



где X – величина варьирующего признака;

 n – объем выборки;

 Σ – знак суммирования.

Для больших выборок среднюю арифметическую удобнее вычислить косвенным методом по формуле:



где А – условное среднее значение нулевого класса;

 р – частоты;

 а – условное отклонение;

 n – объем выборки;

 i – величина классового промежутка.

***Задание.*** Пользуясь вариационным рядом, представленным в таблице 1, составить таблицу 2 для вычисления средней арифметической косвенным методом.

## Распределение вариант по весу Таблица 1

|  |  |
| --- | --- |
| **Границы классов (Wн – Wк)** | **Частоты (р)** |
| 42 – 45 | 1 |
| 46 – 48 | 5 |
| 49– 51 | 12 |
| 52 – 54 | 14 |
| 55 – 57 | 8 |
| 58 – 60 | 6 |
| 61 – 63 | 2 |
| 64 – 67 | 2 |
|  | Σр = n = 50 |

Таблица 2

Рабочая таблица для вычисления средней арифметической

методом условных отклонений

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№ класса** | **Границы классов** **(Wн – Wк)** | **Частоты (р)** | **Условные отклонения (а)** | **Произведение условных отклонений на частоты (ра)** |
| 1 | 42 – 45 | 1 | –3 | –3 |
| 2 | 46 – 48 | 5 | –2 | –10 |
| 3 | 49 – 51 | 12 | –1 | –12 |
| 4 | 52 – 54 | 14 | 0 | 0 |
| 5 | 55 – 57 | 8 | 1 | 8 |
| 6 | 58 – 60 | 6 | 2 | 12 |
| 7 | 61 – 63 | 2 | 3 | 6 |
| 8 | 64 – 67 | 2 | 4 | 8 |
|  | Σр = n = 50 |  |  | Σра = 9 |

Для вычисления средней арифметической необходимо:

1. Найти в построенном вариационном ряду условный средний класс. В качестве условного среднего класса рекомендуется брать класс, который занимает центральное место в данном вариационном ряду и имеет наибольшее по сравнению с другими классами значение частот (р). В нашем примере условным средним классом будет четвертый класс с наибольшей встречаемостью вариант (р = 14) и варьированием веса в пределах 52 – 54 кг.
2. Выделить условный средний класс линиями и принять за нулевой.
3. Вычислить условное среднее значение нулевого класса. Его обозначают буквой **А**.

В нашем примере

1. Определить условное отклонение (**а**) каждого класса от нулевого путем вычитания порядкового номера нулевого класса от порядкового номера других классов. Вверх от класса, принятого за условный нулевой, получим натуральный ряд отрицательных чисел (–1, –2, –3 и т.д.), вниз – натуральный ряд положительных числе (+1, +2, +3 и т.д. в зависимости от класса).
2. Найти произведение частоты на условное отклонение для каждого класса (**ра**) и заполнить графу.
3. Найти сумму частот (Σр = n = 50).
4. Вычислить сумму произведений частот на условное отклонение. Она равна:

Σра = –25+34 =9.

1. Вычислить среднее арифметическое по формуле:

где А – условное среднее значение нулевого класса;

 i – величина классового промежутка.

Таким образом, средний вес равен 53,5 кг.

## Показатели изменчивости

Средние величины характеризуют всю выборку в целом. Но основное свойство ее членов – свойство изменяться от особи к особи – остается при этом нераскрытым.

 Для суждения о степени изменчивости или вариабельности признаков в биометрии наиболее часто используются следующие показатели:

* лимит или размах изменчивости;
* среднее квадратическое или стандартное отклонение;
* коэффициент вариации или изменчивости.

Лимит или разница между максимальным и минимальным значениями признака в выборке является наиболее простым, но и наиболее точным способом количественного выражения степени изменчивости этого признака.

Например, вес спортсменов max = 67 кг, min = 42, lim = 67–42 = 25 кг.

Основным показателем изменчивости является среднее квадратическое отклонение. Среднее квадратическое или стандартное отклонение – это статистическая величина, которая показывает, насколько признак, присущий данному варианту, отклоняется от средней арифметической для данной выборки.

 Среднее квадратическое отклонение обозначают либо греческой буквой S, либо сигма. Для малых выборок среднее квадратическое отклонение вычисляют по формуле :

 Вычисление среднего квадратического отклонения для малых выборок производят в следующем порядке:

1. Находят отклонение каждого варианта от средней арифметической для данной выборки, т.е. устанавливают центральные отклонения.
2. Центральные отклонения возводят в квадрат, чтобы избавиться от отрицательных чисел.
3. Находят сумму квадратов.

**Пример.** Представлена совокупность, состоящая из 5 особей. Все они имеют одинаковый возраст и относятся к одной группе. Нужно вычислить среднюю длину их тела и среднее квадратическое отклонение этого признака.

1. Составим простой вариационный ряд (табл.3)

Таблица 3

|  |  |  |
| --- | --- | --- |
| Показатели вариационного ряда | Особи№1 №2 №3 №4 №5 | Статистические показатели |
| Варианты ряда (длина тела в см) | 45 40 38 35 32 | Средняя арифметическая Х ср= 38 см |
| Отклонение каждой варианты от средней арифметической X–Xср | +7 +2 0 –3 –6 | Сумма всех отклонений Σ (Х –Хср) = 0 |
| Квадраты отклонений  (X–Xср)2 | 49 4 0 9 36 | Сумма квадратов отклонений Σ (Х –Хср)2 = 98 |

2. Вычислим среднюю арифметическую Х:





1. Вычислим отклонения размеров длины тела от средней арифметической

(Х–Хср) и полученные данные проставим в таблицу.

1. Так как сумма отклонений всегда равна нулю Σ (Х–Хср) = 0, то отклонения следует возвести в квадрат и определить сумму квадратов отклонений. В данном примере они будут равны:

Σ (Х–Хср)2 = 49+4+0 + 9 + 36 = 98;

#### Вычисление среднего квадратического отклонения

**для больших выборок**

***Задание.*** Вычислить среднее квадратическое отклонение (S) для данной группы спортсменов по весу

1. Составить вариационный ряд (табл. 4).
2. Определить частоту (р) значений веса в каждом классе.
3. Найти условные отклонения (а) от условного среднего класса.
4. Найти произведение частоты на условное отклонение (графа 5).
5. Условное отклонение возвести в квадрат (графа 4).
6. Вычислить произведение частоты на квадрат условного отклонения (графа 6).
7. По формуле вычислить среднее квадратическое отклонение:





Таблица 4

## Вычисление среднего квадратического отклонения

|  |  |  |
| --- | --- | --- |
| **Границы классов****(Wн–Wв)** | **Частоты****(р)** | **Условные отклонения** |
| **(а)** | **(а)2** | **ра** | **ра2** |
| 42 –45 | 1 | –3 | 9 | –3 | 9 |
| 46 –48 | 5 | –2 | 4 | –10 | 20 |
| 49 – 51 | 12 | –1 | 1 | –12 | 12 |
| 52 – 54 | 14 | 0 | 0 | 0 | 0 |
| 55 – 57 | 8 | 1 | 1 | 8 | 8 |
| 58 – 60 | 6 | 2 | 4 | 12 | 24 |
| 61 – 63 | 2 | 3 | 9 | 6 | 18 |
| 64 – 67 | 2 | 4 | 16 | 8 | 32 |
| Σр = n = 50 | Σра = 9 |  |  |  | Σра2 = 123 |

 Нужно обратить внимание на то, что S имеет два знака (+ и –). Это свидетельствует об отклонении вариант от средней арифметической как в положительную, так и в отрицательную сторону. Среднее квадратическое отклонение является показателем разнообразия признака. Согласно правилу 3 S почти все варианты должны укладываться в интервал от –3 S до +3 S, если минимальный вариант (min вес ) не ниже Х–3S, а максимальный (max вес) не выше Хср+3 S, то наблюдения ведутся над однородной генеральной совокупностью.

 В нашем примере Хср + 3 S = 53,5 + 14,04 = 67,54

 Хср – 3 S = 53,5 – 14,04 = 39,46

Как видим, минимальный вариант 42 не ниже Х – 3 S , т.е. 39,46, а максимальный вариант 67 не превышает Х=3 S, т.е. 67,64. Таким образом, выборка однородна, и изучаемые особи относятся к одному вариационному ряду. Среднее квадратическое отклонение выражается в тех же единицах, которыми измеряется признак, т.е. является поименованной величиной.

**Коэффициент изменчивости (вариации)**

 Основное достоинство среднего квадратического отклонения заключается в том, что оно дает полную количественную характеристику изменчивости изучаемого показателя. Однако сравнить изменчивость двух групп с разными средними значениями изучаемого признака и, тем более, изменчивость разных признаков с помощью данного показателя нельзя. Вот здесь на помощь и приходит следующий показатель изменчивости – коэффициент изменчивости или вариации.

 Коэффициент изменчивости характеризует изменчивость в относительных величинах. Это отношение среднего квадратического отклонения к средней арифметической для данной выборки, выраженное в процентах. Коэффициент изменчивости определяется по формуле:



где Cv – коэффициент изменчивости;

S – среднее квадратическое отклонение;

Хcp – средняя арифметическая.

Вычислить Cv для изучаемых Вами признаков. В нашем примере:

Коэффициент изменчивости позволяет сравнивать степень изменчивости разных признаков. Чем коэффициент изменчивости выше, тем общая изменчивость признака тоже выше. Низкие коэффициенты изменчивости указывают на генетическую однородность популяции по данным показателям, высокие коэффициенты изменчивости свидетельствуют о ее неоднородности.

Ориентировочно считают, что если Cv< 5% – изменчивость низкая, Cv от 5 до 10% – средняя, Cv >10% – высокая. Максимальное значение коэффициента изменчивости обычно не превышает 30%.

## Нормированное отклонение

 Нормированное отклонение – это показатель, характеризующий отдельную варианту или группу вариант. Обозначается буквой **Н.**

 Нормированное отклонение – это величина, которая указывает, на сколько долей среднего квадратического отклонения каждый конкретный член совокупности отклоняется от средней арифметической. Вычисляется он по формуле:

где Н – нормированное отклонение;

 Хср – средняя арифметическая;

 S – среднее квадратическое отклонение.

 Как и коэффициент изменчивости, нормированное отклонение – величина относительная. Каждая варианта характеризуется определенным значением **Н**. Если **Н** какой-либо варианты равно +1, значит эта варианта больше Х на 1. Чем больше значение **Н**, тем дальше от средней арифметической отстоит данная особь.

## Ошибка статистических величин

 Для изучения изменчивости того или иного признака берут не всех представителей, а только часть их (выборочную совокупность или выборку). В каждом конкретном случае в выборку могут попасть особи, имеющие несколько более высокие или более низкие значения признака, поэтому вычисленные значения биометрических величин будут отражать свойства генеральной совокупности с определенными ошибками. Эти ошибки не могут быть устранены при самой тщательной организации исследований, но их можно учесть. Они получили название ошибок репрезентативности или выборочности.
 Ошибки статистических показателей будут тем больше, чем выше изменчивость признака и чем меньше объем выборки.

Ошибки статистических показателей обозначаются буквой **m.** Чтобы различать, к какому показателю относится ошибка, рядом с условным ее обознчением подстрочно приписывается обозначение данного показателя.

***Например.*** mx – ошибка средней арифметической,

 mS – ошибка среднего квадратического отклонения,

mcv – ошибка коэффициента изменчивости.

 Все ошибки измеряются в тех же единицах, что и сами показатели. Ошибки статистических показателей вычисляются по формулам:

где mx – ошибка средней арифметической,

S – среднее квадратическое отклонение,

n – объем выборки



где mS – ошибка среднего квадратического отклонения,

 S – среднее квадратическое отклонение,

n – объем выборки.

где mcv – ошибка коэффициента изменчивости,

 Сcv – коэффициент измечивости,

 n – объем выборки.

Ошибки статистических показателей позволяют уточнить границы, в которых находится фактическое значение данных показателей. Такими границами считается интервал, равный промежутку: показатель ±2 ошибки.

В нашем примере



2mx = ± 1,3 X = 53,5 ± 1,3 кг.

Вычислить mср ,mS, mcv для изучаемых Вами признаков.

**Критерий достоверности**

**и достоверность разности между средними**

 Критерий достоверности позволяет определить, насколько велика допущенная в опыте ошибка. Его обозначают буквой **t** и вычисляют по формуле:

Если критерий достоверности больше 3 (t>3), то данные опыта достоверны, ошибка составляет около 5%. Если критерий достоверности меньше 3 (t<3),то полученным данным верить нельзя.

 Критерий достоверности зависит от размаха изменчивости и от числа наблюдений. Если t<3, то нужно увеличить выборку, взять для наблюдений больше особей, проверить, нет ли случайных значений вариант. В нашем примере

Полученное число больше 3, значит данные достоверны.

 При сопоставлении данных двух изучаемых групп можно вычислить достоверность разницы данного показателя между ними. Для этого вычисляют критерий достоверности различий. Этот показатель предложил В.С.Госсет, публиковавший свои работы под псевдонимом Стьюдент, отсюда и название – критерий Стьюдента. Критерий Стьюдента **td** вычисляют по формуле:

где td– показатель достоверности разности,

X ср1 – X ср2 – разность между средними арифметическими двух сравниваемых групп (от большего значения Хср вычисляется меньшее),

Разность притнято считать достоверной, если td факт >td табл.

Табличное значение td (t) выбирается по таблице, которая помещается в каждом учебнике по биометрии. Под числом степеней свободы **Y** (**ню**) понимается число наблюдений, уменьшенное на число ограничений.

***Например:*** n1= 50, n2= 50, Y = (n1 –1)+ (n2 – 1)= (50–1) + (50 – 1) = 98.

В таблице даются достоверные величины td при трех порогах вероятности (Р): 0,95; 0,99; 0,999 с учетом числа степеней свободы. Оптимально будет выбрать уровень Р 0,95. В месте пересечения строки значения Y и графы уровня вероятности и находится табличное значение td. Так, при значении Y=99 и Р 0,95 td табличное равно 1,98.

Если tdфакт.< tdтабл., то разность принято считать недостоверной, т.к. при этом нельзя сказать, какая из двух сравниваемых генеральных средних имеет лучшие показатели и сохранится ли подобная зависимость в других исследованиях или опытах.

Например, сравнивая по весу 2 группы детей, имеем:

Х1 = 23 кг (23,0) mхср1 = 3,1 кг,

Х2 = 24 кг (24,0) mхср2 = 2,0 кг,

Таким образом, tdфакт = 0,27< tdтабл. Значит, вес детей в 2-х группах различается недостоверно, т.е. нельзя сказать, какая группа имеет лучшие показатели.

## Построение вариантной кривой

Графическое выражение изменчивости признака, отражающее как размах вариации, так и частоту встречаемости отдельных вариант выполняются в виде вариационной кривой. Вариационная кривая (привести рисунок) строится при изучении количественных признаков, значение которых может быть выражено целым и дробным числом. На графике откладываются значения Хср, а также интервала Х ± S; Х ± 2S; Х ± 3S, а также Mo и Ме.

Мода (Мо) и медиана (Ме) являются дополнительными характеристиками среднего значения варьирующего признака в совокупности.

Мода показывает, какая величина варианта (Хмо)данного признака чаще всего встречается в совокупности.

Медиана указывает на то, какой вариант расположен в середине (центре) вариационного ряда, он делит совокупность на две равные части: с уменьшающимися и увеличивающимися значениями Х от медианы.

Использование Мо и МЕ особенно удобно для сопоставления совокупностей по качественным признакам. Например, модальный цвет кожи у европеоидной расы – белый, модальное число пальцев на руке – 5 и т.д.

В нормальном распределении величины Хср, Мо и Ме совпадают.

При изучении же дискретных количественных признаков (количество детенышей в помете, число индивидуумов с доминантным признаком, количество зубов и т.д.) и качественных признаков – строится гистограмма. Следует помнить, что при изучении изменчивости качественных признаков (цвет глаз, волос, окраска венчика и др.) устанавливается частота в абсолютных значениях и в % в каждом классе изменчивости и вычерчивается гистограмма, а другие показатели изменчивости не вычисляются.

***Задание:***

Вычислить показатели изменчивости качественного и количественного признака по заданию преподавателя. Оформить работу в виде индивидуального задания. Объяснить значение всех вычисленных показателей изменчивости.